Monopole ionization and the transition from weak to strong coupling in gauge theories

Daniel R. Stump
Physics Department, Indiana University, Bloomington, Indiana 47405
(Received 29 April 1980)

The influence of long-range gauge-field configurations on the Wilson loop integral W is discussed. It is shown that W can in principle be found from effective two- or three-dimensional models. The long-range fields considered are in two dimensions a vortex field, and in three dimensions a magnetic monopole field; the connection between these is that a monopole in three dimensions creates a vortex in two dimensions. Models of the statistical distribution of vortices or monopoles are invented and used to study the effect of these fields on W. It is suggested that the transition from weak to strong coupling might be due to a topological order-disorder phase transition in which monopoles are liberated. The XY model is discussed as an analog system with such a phase transition.

I. INTRODUCTION

A quantity of theoretical interest in quantum chromodynamics (QCD) is the quark-antiquark static potential $V(R)$, which is the potential energy of a pair of static quarks in a color-singlet combination separated by the distance R. It is thought that $V(R)$ increases to infinity in the limit $R\to\infty$, for the pure gauge theory, and that this divergence of $V(R)$ is a sign that quark confinement occurs in QCD.

In recent papers[1-3] I have studied the instantaneous Coulomb potential of a static quark pair, for the pure gauge theory with gauge group SU(2), as a model of the static potential $V(R)$. The instantaneous Coulomb potential depends on the distribution of vacuum fluctuations of the gauge fields $A_a^i(\vec{x})$. It is shown in Ref. 3 that coherent long-range vacuum gauge fields $A_a^i(\vec{x})$, i.e., fields that decrease as $|\vec{x}|^{-1}$ as $|\vec{x}|\to\infty$, affect the R dependence of the instantaneous Coulomb potential at large R; the specific long-range field considered is the Wu-Yang monopole field.[4] The observation that this field changes the asymptotic form of the instantaneous Coulomb potential leads to the suggestion that quark confinement might result from the influence of such vacuum fluctuations.[5]

The idea that the vacuum state of the SU(2) gauge theory might be described as a superposition of monopole fields with only small fluctuations away from the monopole fields has been discussed by Mandelstam.[6]

The instantaneous Coulomb interaction is the origin of asymptotic freedom in the Coulomb-gauge formulation of the theory,[3,7] and is therefore a natural quantity to examine for a sign of confinement. However, the instantaneous Coulomb potential is not fundamental because it is not gauge invariant. Thus, it is at best a qualitative model of $V(R)$.

The gauge-invariant potential $V(R)$ can be obtained from the Wilson loop-integral formula.[8] Let Σ be a finite two-dimensional surface in Euclidean four-space with boundary $\partial\Sigma$ a simple closed curve; for simplicity let Σ be planar. The Wilson loop integral W is the vacuum expectation value

$$W=\left\langle 0\left|\tfrac{1}{2}\mathrm{Tr}P\exp\left(ig\oint_{\partial\Sigma}ds_\mu A_a^\mu t_a\right)\right|0\right\rangle, \tag{1.1}$$

where t_a are the matrix generators of the gauge group SU(2) in the fundamental representation and g is the coupling constant. The symbol P indicates path ordering with respect to multiplication of SU(2) matrices and time ordering with respect to multiplication of gauge-field operators $A_a^\mu(x)$. The quantity W can be written as a Euclidean path integral as

$$W=Z^{-1}\int dA_a^\mu(x)e^{-S(A)}\tfrac{1}{2}\mathrm{Tr}P\times\exp\left(ig\oint_{\partial\Sigma}ds_\mu A_a^\mu t_a\right), \tag{1.2}$$

where $S(A)$ is the Euclidean action

$$S(A)=\int d^4x\,\tfrac{1}{4}F_a^{\mu\nu}F_a^{\mu\nu} \tag{1.3}$$

and gauge-fixing and Faddeev-Popov ghost terms have been suppressed for simplicity of notation; the normalization factor Z is

$$Z=\int dA_a^\mu(x)e^{-S(A)}. \tag{1.4}$$

If the loop $\partial\Sigma$ lies in the (x_4,x_1) plane, where x_4 refers to Euclidean time, and has lengths T and R in the x_4 and x_1 directions, then it can be argued[8,9] that in the limit $T\to\infty$ with R fixed

$$W\sim c^2e^{-TV(R)}, \tag{1.5}$$

where c is independent of T and $V(R)$ is the energy of the lowest-energy state containing a $q\bar{q}$ pair with separation R. If $V(R)$ is linearly confining, i.e., $V(R)\sim\gamma R$ as $R\to\infty$, then

$$W\sim c^2e^{-\gamma A(\Sigma)}, \tag{1.6}$$

where $A(\Sigma)=RT$ is the area of the surface Σ. The constant γ is called the string tension. Thus, one formulation of the question of quark confinement in a gauge theory is simply whether the loop integral W obeys the area law Eq. (1.6) as the size of Σ tends to infinity. If so, the theory is said to be confining; if not, i.e., if instead W obeys a perimeter law $W\sim\exp[-\lambda L(\partial\Sigma)]$, where $L(\partial\Sigma)$ is the length of $\partial\Sigma$, the theory is said to be nonconfining.

The gauge theory should properly be defined as a lattice theory[8] to avoid ultraviolet singularities in the quantity W.

By rotational invariance of the Euclidean theory, W is independent of the orientation of the surface Σ. In what follows, Σ will be taken to lie in the (x_1, x_2) plane. Then Eq. (1.1) shows that W is determined by the distribution of time-independent configurations of the gauge fields in the vacuum state. On the other hand, the path-integral formulation Eq. (1.2) can be thought of as a kind of statistical mechanics problem: The variables are field configurations $A_a^\mu(x)$ with energy $H=\beta^{-1}S(A)$ and partition function Z; in this language W is the ensemble average of the loop integral $\exp(i\oint ds\cdot A)$. The inverse temperature β need not be introduced explicitly but can be absorbed into the definition of the coupling constant g.

The purpose of this paper is to examine the influence of coherent long-range gauge fields, such as monopole fields, on the loop integral W. To be specific, *simple models* of the contribution of such fields to functional integrals that define Z and W will be invented and used to study the question: Under what circumstances does W decrease exponentially as the area of Σ (area law) as in Eq. (1.6), and under what circumstances as the length of the boundary curve $\partial\Sigma$ (perimeter law)?

The study of the influence of monopole fields on the loop integral W will involve two steps. First it will be shown that to calculate the quantity W the theory can in principle be reduced to an effective two-dimensional (2D) theory. The coherent long-range fields of interest in the 2D theory are vortex configurations. Simple statistical-mechanics *models* of the distribution of vortices will be used to study their effect on W. Second, three-dimensional (3D) monopole fields in the vacuum state and their effect on W will be considered, again by looking at statistical-mechanics models of the monopole distribution. The connection between these two approaches is that 3D monopole fields create 2D vortex fields.

The 2D vortex models and 3D monopole models that are to be considered are obtained by replacing the functional integrals that define Z and W by integrals over just the collective coordinates of the vortex or monopole configurations. The idea behind this replacement is suggested by instanton calculations.[10,11] The assumption is that the functional integrals over field configurations are dominated by two kinds of fields: (1) long-range topological configurations, vortices (2D) or monopoles (3D), which in a lattice gauge theory could be treated in a continuum approximation because they spread over many lattice sites; and (2) typical quantum fluctuations away from the topological configurations, which are small and hence could be treated in Gaussian approximation. The long-range fields are assumed to have the dominant effect on the large-Σ behavior of W, so only the vortex or monopole sector need be considered. This replacement reduces the problem to a statistical-mechanics model of a gas of vortices or monopoles.

A system in which this separation of fields into topological configurations and Gaussian fluctuations is quite explicit is the XY model in the treatment of Kosterlitz and Thouless.[12] The analogy between the spin vortices in their theory of the XY model and merons[13] in QCD is described in Ref. 14. The XY model will be discussed further in Sec. V in a way that emphasizes analogies with the discussions in the present paper.

Models of the monopole sector of the theory that are particularly interesting are those that have two phases, a disordered phase with free monopoles and an ordered phase in which monopoles are bound in dipole pairs. The phase transition separating these phases is the monopole ionization referred to in the title of this paper. It will be shown that an ionized phase is associated with an area-law behavior of W, and a dipole-pair phase with a perimeter law.

It is believed that Abelian lattice gauge theories have two phases, a disordered confining phase at large g^2 and an ordered nonconfining phase at small g^2, separated by a phase transition at some critical point g_c^2.[15] It has been suggested[16] that this phase transition is caused by the appearance of Dirac monopole vacuum fields at $g^2\geq g_c^2$ that are not present for $g^2<g_c^2$. It is this transition from strong to weak coupling, from area-law behavior of W to perimeter law, that is referred to in the title of this paper.

Non-Abelian lattice gauge theories are expected *not* to have a phase transition to a nonconfining phase at small g^2. However, the string tension γ of these theories must have a different g^2 dependence for large g^2 and small g^2, and there is evidence that the transition from strong- to weak-coupling behavior occurs quite abruptly.[17,18] It is possible that this transition is also associated with monopole vacuum fields, and that in the non-

Abelian theory the monopole ionization occurs abruptly but continuously. In a non-Abelian gauge theory, monopole vacuum fields result from the contribution of meron paths to the path integral Eq. (1.2). The possible role of merons in the transition from strong to weak coupling will be discussed briefly in Sec. IV.

The ionization transition from dipole pairs to free monopoles can be considered a topological order-disorder transition, just as the phase transition in the XY model involves topological order.[12] The role of topological symmetry breakdown in quark confinement has been considered also by Samuel.[19]

A different explanation of the transition from strong to weak coupling in non-Abelian gauge theories, involving only effects of instantons, has been proposed by Callan *et al.*[20]

The approach to the problem of understanding W that is taken in this paper is frankly heuristic. Some of the remarks could perhaps be formulated more rigorously in the context of a four-dimensional (4D) lattice gauge theory. The aim of the discussion here is not to *derive* the role of monopole fields in the theory but just to illustrate how they might influence W with explicit models.

The outline of the paper is as follows. In Sec. II the effective 2D theory that determines W is described, and the influence of 2D vortices on W is examined. In Sec. III, 3D monopoles are discussed in their effect on W and their role in creating vortices in the 2D theory. In addition, a brief discussion of dual parameters, such as the 't Hooft loop operator,[21] is given. Section IV concerns the origin of 3D monopoles in gauge theories, and is largely a repetition of remarks from Ref. 16. Finally, in Sec. V the XY-model analogy is developed.

II. TWO-DIMENSIONAL VORTICES AND THE WILSON LOOP INTEGRAL

In this section it will be shown that to calculate the quantity W the theory reduces to an effective two-dimensional theory. The influence of 2D vortices on W will be explored by considering simple models of the distribution of vortices.

The full theory described by the path integral in Eq. (1.2) is four-dimensional: x is a point in Euclidean four-space and $A_a^\mu(x)$ is a four-vector field. However, the Wilson loop operator W depends only on the value of the field on the two-dimensional subspace Σ. In particular, if Σ is taken to lie in the (1, 2) plane then the expectation value W reduces to the expectation value in an effective 2D model on the (12) plane. Let the reduced probability distribution $P_2(\varphi_a^\alpha)$ be defined by

$$P_2(\varphi_a^\alpha)=\int dA_a^\mu(x)e^{-S(A)} \times\prod_{x_\perp}\prod_{\substack{a=1,2,3\\ \alpha=1,2}}\delta[\varphi_a^\alpha(x_\perp)-A_a^\alpha(x_\perp,0,0)], \tag{2.1}$$

where $\varphi_a^\alpha(x_\perp)$ is a two-vector field ($\alpha=1,2$) in a 2D space with point $x_\perp=(x_1,x_2)$; that is, $P_2(\varphi_a^\alpha)$ is the reduced probability distribution of fields on the 2D subspace, the (12) plane. The loop integral W is determined from $P_2(\varphi_a^\alpha)$ by

$$W=Z^{-1}\int d\varphi_a^\alpha(x)P_2(\varphi_a^\alpha)\tfrac{1}{2}\,\mathrm{Tr}P\exp\left(ig\oint_{\partial\Sigma}ds_\alpha\varphi_a^\alpha t_a\right) \tag{2.2}$$

because Σ lies in the (12) plane; the partition function Z is

$$Z=\int d\varphi_a^\alpha(x)P_2(\varphi_a^\alpha). \tag{2.3}$$

An effective action $s_2(\varphi)$ can be defined by $s_2(\varphi)=-\ln P_2(\varphi)$, and W can be thought of as a correlation function of the effective 2D theory.

It would be necessary to solve the full 4D theory in order to determine $P_2(\varphi)$. That is intractable. However, it might be of some value to consider some simple models of $P_2(\varphi)$ as a first step toward learning how the area-law behavior of W in Eq. (1.6) can be produced. To be specific, the influence of coherent long-range fields will be considered.

The long-range 2D field configurations to be considered are vortices in the field $\varphi_a^\alpha(x)$. For simplicity, vortex configurations for an Abelian gauge theory will be discussed first. The point vortex field, centered at $x=0$, is

$$\varphi_v^\alpha(x)=-\frac{1}{2g}\epsilon_{\alpha\beta}\frac{x_\beta}{x^2}=\frac{1}{2g}\partial_\alpha\theta, \tag{2.4}$$

where $\theta=\arctan x_2/x_1$. The loop integral of that field is

$$g\oint_{\partial\Sigma}ds_\alpha\varphi_v^\alpha=\pi\theta_\Sigma(0), \tag{2.5}$$

where the function $\theta_\Sigma(x)$ is defined as

$$\theta_\Sigma(x)=\begin{cases}1 & \text{if } x\in\Sigma\\ 0 & \text{if } x\notin\Sigma.\end{cases} \tag{2.6}$$

The quantity that appears in W is $\exp(ig\oint ds\cdot\varphi)$, which has the absolute value 1 and is equal to +1 for $\varphi=0$ and -1 for $\varphi=\varphi_v$ if $\partial\Sigma$ surrounds the vortex center. The special significance of a vortex field is first that it obeys a flux quantization condition, namely, that the total flux of $\nabla\times\vec{\varphi}$ is π/g, that implies $W=-1$; and second that it is of long range, i.e., is of order $|x|^{-1}$ as $|x|\to\infty$,

and thus contributes to the line integral around $\partial\Sigma$ for arbitrarily large Σ.

An antivortex field, with opposite vorticity, is simply $-\varphi_v^\alpha$. Multivortex fields can be defined as superpositions of vortices; the general multivortex field is

$$\overline{\varphi}^\alpha(x)=\sum_{i=1}^{n}\varphi_v^\alpha(x-x_i)-\sum_{j=1}^{n'}\varphi_v^\alpha(x-y_j), \tag{2.7}$$

where x_i and y_j are the centers of vorticity of n vortices and n' antivortices. The loop integral of $\overline{\varphi}^\alpha(x)$ gives

$$\exp\left(ig\oint_{\partial\Sigma} ds\cdot\overline{\varphi}\right)=\prod_{i,j} e^{i\pi\theta_\Sigma(x_i)}e^{-i\pi\theta_\Sigma(y_j)}. \tag{2.8}$$

It is possible to construct a similar point vortex field for the SU(2) gauge theory, i.e., a long-range field obeying a flux quantization condition such that $W=-1$ if $\partial\Sigma$ surrounds the center of vorticity. The flux quantization condition for a non-Abelian vortex can be described as follows: the line-integral operator $P\exp(i\oint ds_\alpha\cdot\varphi^\alpha)$ is the operator that describes parallel transport in the field $\varphi_a^\alpha t_a$ of vectors v of the basis of the fundamental representation of SU(2); flux quantization of the vortex field φ_a^α means that parallel transport of v around $\partial\Sigma$ carries v to $-v$ if $\partial\Sigma$ surrounds the vortex center. A specific example of such a configuration is an embedding of $\varphi_v^\alpha(x)$ into the non-Abelian theory

$$\varphi_{va}^\alpha(x)=2n_a\varphi_v^\alpha(x), \tag{2.9}$$

where n_a is any constant unit vector, $n_a^2=1$. This field satisfies the topological flux-quantization condition $P\exp(ig\oint ds\cdot\varphi_a t_a)=-1$ if t_a are the generators of the fundamental representation of SU(2). It is interesting to note that $W=+1$ if t_a are adjoint-representation generators.

It is not known to me how to generalize the multivortex configuration in Eq. (2.7) to a non-Abelian theory. Simple superposition of single vortex fields does not make sense in a nonlinear theory. A multivortex configuration would presumably have additional degrees of freedom, besides just the positions of vortex centers, associated with relative orientations of the vortices in SU(2) space [i.e., the choice of n_a in Eq. (2.9)]. In what follows I shall assume that such a generalization exists but that the additional degrees of freedom are not important for understanding the influence of vortices on the quantity W. That is, Eq. (2.8) will be assumed to be appropriate to describe the loop integral of a multivortex field even in a non-Abelian theory.[22]

In Sec. III it will be argued that vacuum fluctuations of the form of monopole fields create vortex-like configurations in the 2D theory.

In the remainder of Sec. II the influence of 2D vortices on the loop integral W will be studied by considering some simple models of the vortex contribution to the 2D functional integrals in Eqs. (2.2) and (2.3). These models are obtained by replacing the functional integrals by integrals over just the coordinates of the vortex centers and summing over all vortex configurations. That is, the quantities Z and W are replaced by

$$Z=\sum_{n,n'=0}^{\infty}\frac{1}{n!n'!}\int\prod_{i=1}^{n}d^2x_i\prod_{j=1}^{n'}d^2y_jP(x,y) \tag{2.10}$$

and

$$W=Z^{-1}\sum_{n,n'=0}^{\infty}\frac{1}{n!n'!}\int\prod_{i,j}d^2x_id^2y_jP(x,y)\times e^{i\pi\theta_\Sigma(x_i)}e^{-i\pi\theta_\Sigma(y_j)}, \tag{2.11}$$

where $P(x,y)$ is the probability distribution of configurations with vortex and antivortex positions x_i and y_j; the factor $(n!n'!)^{-1}$ is the Boltzmann counting factor that appears because vortex configurations that differ only by exchange of x_i's or y_j's are identical. The integrand of Eq. (2.11) for W is the loop operator evaluated for the multivortex field $\overline{\varphi}$ in Eq. (2.8), and it is assumed that this form would also be appropriate for a non-Abelian multivortex field.

The idea of using these expressions as models of Z and W is suggested by instanton calculations, in the manner described in Sec. I: The assumption is that the partition sum Z is dominated by fields $\varphi^\alpha(x)$ that differ from the multivortex field $\overline{\varphi}^\alpha(x)$ only by typical perturbative fluctuations, and that these small fluctuations do not have a major effect on W. This assumption would lead to the replacement of the loop operator by its value for the field $\overline{\varphi}^\alpha$ and to integrals over the collective coordinates x_i and y_j as in Eqs. (2.10) and (2.11); the probability distribution $P(x,y)$ would be determined by the volume in field-configuration space of the small fluctuations of φ^α away from $\overline{\varphi}^\alpha$.

The replacement of the full average over field configurations by an average over vortex configurations reduces the problem to the statistical mechanics of a 2D gas of vortices with probability distribution $P(x,y)$. The quantity Z, for example, is the partition function of the grand canonical ensemble. Determination of the distribution $P(x,y)$ would require evaluation of the path integral $\int d\varphi^\alpha(x)$. That is not attempted here. Rather, some simple models are considered in which $P(x,y)$ is chosen *ad hoc*. It is hoped that these models illustrate how the vortices might influence W in the full theory.

The first model to be considered is a perfect gas of vortices and antivortices. That is, $P(x,y)$ is just taken to be

$$P(x,y)=\xi^{n+n'}. \tag{2.12}$$

Since P is independent of the positions x_i and y_j, the vortices do not interact. The parameter ξ, which has units (length)$^{-2}$, is the density of vortices. The partition function is

$$Z=e^{2\xi\Omega}, \tag{2.13}$$

where $\Omega=\int d^2x$ is the volume of the 2D space, kept finite for the moment. The loop integral is

$$W=e^{-2\xi\Omega}\sum_{n,n'}\frac{\xi^{n+n'}}{n!n'!}\left(\int d^2x\, e^{i\pi\theta_\Sigma(x)}\right)^n\left(\int d^2y\, e^{-i\pi\theta_\Sigma(y)}\right)^{n'}$$
$$=e^{-A(\Sigma)\xi}, \tag{2.14}$$

where $A(\Sigma)=\int d^2x\,\theta_\Sigma(x)$ is the area of the region Σ. Thus a perfect gas of vortices produces the area-law behavior associated with a linearly confining potential with string tension $\gamma=4\xi$.

The next model is a perfect gas of vortex-antivortex bound pairs, in which the $v\bar{v}$ separation length of each pair is fixed at l. That is, $P(x,y)$ vanishes unless the centers x_i and y_j of vorticity are paired with $|x-y|=l$ for each pair; and if there are n pairs then P is

$$P=\xi'^n, \tag{2.15}$$

where ξ' is the density of pairs. Again P is independent of the positions of the $v\bar{v}$ molecules so they are noninteracting. The degrees of freedom of each molecule are the position z of the center of mass and angular orientation θ with respect to the x_1 axis. The partition function of this system is

$$Z=\sum_{n=0}^{\infty}\frac{\xi'^n}{n!}\prod_{i=1}^{n}\int d^2z_i\int\frac{d\theta_i}{2\pi}=e^{\xi'\Omega}. \tag{2.16}$$

The loop integral is

$$W=e^{-\xi'\Omega}\sum_n\frac{\xi'^n}{n!}\left[\int d^2z\int\frac{d\theta}{2\pi}e^{i\pi\theta_\Sigma(x)}e^{-i\pi\theta_\Sigma(y)}\right]^n$$
$$=\exp\left\{-\xi'\int d^2z\int\frac{d\theta}{2\pi}\{1-\cos\pi[\theta_\Sigma(x)-\theta_\Sigma(y)]\}\right\}, \tag{2.17}$$

where x and y are the v and $\bar{v}$ positions corresponding to the $(v\bar{v})$ coordinates z, θ; that is,

$$x=z+\tfrac{1}{2}\rho,$$
$$y=z-\tfrac{1}{2}\rho, \tag{2.18}$$
$$\rho=(l\cos\theta, l\sin\theta).$$

Finally, W can be written

$$W=\exp\left[-2\xi'\int d^2z\int\frac{d\theta}{2\pi}F_\Sigma(x,y)\right], \tag{2.19}$$

where the function $F_\Sigma(x,y)$ is equal to 1 if one of x and y is in Σ, and is equal to zero otherwise. The integral in Eq. (2.19) cannot be evaluated exactly for an arbitrary region Σ, but it is not difficult to obtain the limiting form when the size of Σ is large compared to l:

$$W\sim\exp\left[-L(\partial\Sigma)\frac{4}{\pi}\xi' l\right]\quad\text{for}\ \frac{l}{L(\partial\Sigma)}\ll 1, \tag{2.20}$$

where $L(\partial\Sigma)$ is the length of the curve $\partial\Sigma$. Thus, a perfect gas of $v\bar{v}$ pairs gives the perimeter-law behavior associated with a nonconfining gauge theory. The origin of the perimeter law is that the $v\bar{v}$ configurations that affect W significantly are those for which only one of the members of the pair lies in Σ; for Σ large compared to l, only pairs with position near $\partial\Sigma$ are relevant.

The loop integral W provides a measure of the amount of disorder in the system: a small value of W indicates a large amount of disorder. For example, in either of the above models, an increase in the density ξ of vortices or $v\bar{v}$ pairs results in a decrease of W. Furthermore, W decreases as the size of the loop $\partial\Sigma$ increases and the rate of this decrease of W is also sensitive to the amount of disorder: in either of the above models, the rate of decrease of W increases as the density ξ increases. The distinction between an area-law decrease and a perimeter-law decrease is a more subtle measure of the disorder in the system. This distinction involves more than just the *amount* of disorder: In the perfect gas of $v\bar{v}$ pairs, W obeys a perimeter law no matter how large ξ' and l are, whereas in the perfect gas of vortices W obeys an area law no matter how small ξ is. In these models the distinction between area law and perimeter law measures the special disorder created by the coherent long-range vortex fields, i.e., the *topological disorder*, a term used in Ref. 12.

The final 2D vortex model, which will be discussed presently, has two phases, one with free vortices and one with only bound $v\bar{v}$ pairs; there is a phase transition that can be called vortex ionization. The model is a two-dimensional Coulomb gas, or plasma, of vortices and antivortices. Before describing that familiar system[12,23] some general remarks will be made concerning the possible relevance of such a phase transition to the gauge theory.

A transition from free vortices to $v\bar{v}$ pairs would imply a transition from a confining phase to a nonconfining phase. The vortex ionization in the 2D slice of the theory would reflect some other process in the full 4D theory, such as a sudden increase in the contribution to the 4D path integral of some special topological field configurations

that create vortices in the 2D subspace. In a non-Abelian gauge theory, meron ionization in the 4D theory might serve this purpose. A model of this kind of transition is the vortex ionization transition in the XY model.

An ionization transition like that described in the previous paragraph would manifest itself not only by the change in the asymptotic form of the correlation function W but also by changes in ordinary thermodynamic functions of the system. The nature of these functions can be determined from the results of recent Monte Carlo studies of an SU(2) lattice gauge theory.[18,24,25] In particular, it is interesting to compare the results of those calculations to a similar Monte Carlo study of the XY model.[26] The transition of the XY model is reasonably well understood as a vortex ionization effect,[12,27] and comparison of the two Monte Carlo studies might indicate whether an ionization transition is inconsistent with the observed behavior of the thermodynamics of the gauge theory.

The quantity that will be compared in these two systems is the specific heat. Consider a system with partition function Z and inverse temperature β. The free energy F is defined by

$$\beta F = -\frac{1}{\Omega}\ln Z ,$$

where Ω is the volume of the system. The energy per unit volume E is

$$E = \frac{\partial}{\partial\beta}(\beta F) .$$

The thermodynamic entropy S is

$$S = \beta E - \beta F .$$

Finally, the specific heat C can be written in two ways as

$$C = -\beta^2 \frac{\partial E}{\partial \beta} = -\beta \frac{\partial S}{\partial \beta} .$$

The specific heat C measures the rate at which entropy increases as β decreases. It is related to the mean-square fluctuation of the energy by

$$C = \frac{1}{\Omega}\beta^2\left[\frac{1}{Z}\frac{\partial^2 Z}{\partial\beta^2} - \left(\frac{1}{Z}\frac{\partial Z}{\partial\beta}\right)^2\right] .$$

Figure 1 shows the specific heat C as a function of β for the XY model computed by Tobochnik and Chester[26] by Monte Carlo methods, for a lattice of size 60×60. The parameter β is $1/T$ where T is the temperature parameter of Ref. 26. The volume Ω of the system, used in the formula for C, is taken to be the number of lattice sites (=3600). It can be shown that the low-β behavior of C is $C\sim\beta^2$ as $\beta\to0$, and the high-β behavior is $C\to\frac{1}{2}$ as $\beta\to\infty$. There is a broad peak around β

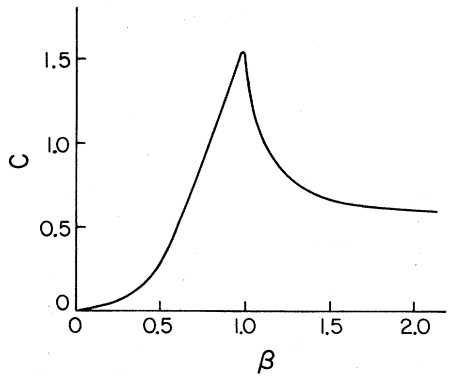

FIG. 1. The specific heat C vs inverse temperature $\beta=1/T$ for the XY model, taken from the Monte Carlo study of Ref. 26. The vortex-ionization transition point is estimated to be at $\beta=1.12$.

$\simeq 1$ but no singularity. The vortex ionization point of the XY model is estimated to be at $\beta_c = 1.12$. The maximum of C reflects the increase in entropy associated with appearance of the new degrees of freedom of the ionized vortices.

Figure 2 shows the specific heat C *vs* β for an SU(2) lattice gauge theory. This curve was obtained from data on the average energy E calculated by Monte Carlo methods. To be specific, Petcher and Weingarten[25] have computed E for a 120 element discrete subgroup of SU(2) for an $8\times8\times8\times8$ lattice; the values of E *vs* β for this system are very similar to those found by Creutz[18] for the continuous group SU(2), at least over the range of β shown in Fig. 2. The specific heat C is obtained from E by differentiation with respect to β. The statistical error in E is sufficiently large so that in order to obtain a valid estimate of C it is necessary to fit a smooth curve to the computed values of E and find C from the smooth curve. In this

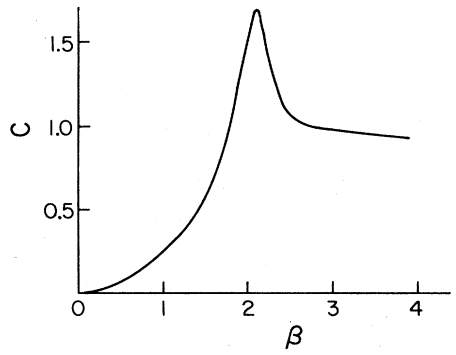

FIG. 2. The specific heat C vs $\beta=4/g^2$ for an SU(2) lattice gauge theory, taken from the Monte Carlo study of Ref. 25. The values of C on the curve are accurate only to within an error of roughly 5 to 10%.

problem the coupling constant g^2 and inverse temperature β are related by $\beta = 4/g^2$. The volume Ω of the system is taken to be the number of plaquettes of the lattice ($=6\times 8^4$). It can be shown that the low-β form of C is $C \sim \frac{1}{4}\beta^2$ as $\beta \to 0$, and the high-β form is $C \to \frac{3}{4}$ as $\beta \to \infty$.

The overall qualitative features of the curve in Fig. 2 are similar to the XY-model curve. The scales of the C and β axes in these figures are not particularly meaningful because of ambiguities in the definitions of β and Ω for the two systems. The scales have been chosen such that the low-β forms of the curves are the same. There is a broad peak in the specific heat of the gauge theory around $\beta \simeq 2$. It should be noted that the specific heat is quite different for gauge theories with phase transitions such as a 4D U(1) gauge theory or a 5D SU(2) gauge theory[18]; there C is singular.

Monte Carlo computations of the string tension γ as a function of β for SU(2) lattice gauge theories have also been carried out.[18,25] It is found that γ is nonzero for all β considered, and that the dependence of γ on β changes abruptly, but continuously, from strong-coupling dependence for small β

$$\gamma \sim -\ln\beta/4 \tag{2.21a}$$

to weak-coupling dependence for large β

$$\gamma \sim \exp\left[-\frac{6\pi^2}{11}(\beta-2)\right]. \tag{2.21b}$$

The transition from strong to weak coupling occurs over precisely the interval in β around $\beta \simeq 2$ where the peak in C is located. Thus the transition from weak to strong coupling is accompanied by an increase in available degrees of freedom as indicated by a peak in $\partial S/\partial\beta$. The similarity of the curves in Figs. 1 and 2 indicates that the behavior of the specific heat does not contradict the idea that this transition might involve some kind of ionization of topological configurations, like the vortex ionization of the XY model.

Finally, a 2D model will be considered that has a transition from a phase with bound $v\bar{v}$ pairs to an ionized vortex phase—the 2D Coulomb gas of vortices. This system is formulated on a 2D lattice with lattice spacing a in order to eliminate ultraviolet singularities. The treatment to be described follows that of Ref. 12.

The partition function is

$$Z = \sum_{n,n'} \frac{1}{n!n'!} \sum_{\substack{x_1,\ldots,x_n \\ y_1,\ldots,y_{n'}}} P(x,y), \tag{2.22}$$

where x_i $(i=1,\ldots,n)$ and y_j $(j=1,\ldots,n')$ are vortex and antivortex lattice positions. The probability distribution $P(x,y)$ is

$$P(x,y) = (\xi a^2)^{n+n'} e^{-\beta H(x,y)}, \tag{2.23}$$

where ξ is a density parameter and the energy H is

$$H = \tfrac{1}{2}\kappa^2 \sum_{\lambda,\lambda'=1}^{n+n'} e_\lambda e_{\lambda'} G(x_\lambda - x_{\lambda'}), \tag{2.24}$$

where κ is the "charge" of the vortex and e_λ and x_λ stand for

$$e_\lambda, x_\lambda = \begin{cases} +1, x_i & \text{for } \lambda = i,\ i=1,\ldots,n, \\ -1, y_j & \text{for } \lambda = n+j,\ j=1,\ldots,n'. \end{cases} \tag{2.25}$$

The potential $G(x-x')$ is the Green's function of the lattice Laplacian

$$-\nabla^2 G(x-x') = \delta^2(x-x') \tag{2.26}$$

with the boundary condition that G vanishes at $x=x'$. The function G can be written

$$G(x) = a^2 \int_{-\pi/a}^{\pi/a} \frac{dp_1}{2\pi}\frac{dp_2}{2\pi} \frac{\cos p\cdot x - 1}{4 - 2\cos p_1 a - 2\cos p_2 a}; \tag{2.27}$$

it is well approximated, even for small x, by its large-x asymptotic form

$$G(x) \simeq -\frac{1}{2\pi}\ln\frac{|x|}{r_0}, \tag{2.28}$$

where

$$r_0 = a\frac{e^{-\gamma}}{2\sqrt{2}} \tag{2.29}$$

and $\gamma = 0.577\cdots$ is Euler's constant. The approximation (2.28) shows the usual logarithmic 2D Coulomb potential. The potential is attractive for a $v\bar{v}$ combination; repulsive for vv or $\bar{v}\bar{v}$. In continuum notation the partition function Z is written

$$Z = \sum_{n,n'} \frac{\xi^{n+n'}}{n!n'!} \int \prod_{i=1}^{n} d^2x_i \prod_{j=1}^{n'} d^2y_j e^{-\beta H}. \tag{2.30}$$

The loop integral W is, in accordance with Eq. (2.11),

$$W = Z^{-1} \sum_{n,n'} \frac{\xi^{n+n'}}{n!n'!} \int \prod_{i,j} d^2x_i d^2y_j e^{-\beta H} e^{i\pi\theta_\Sigma(x_i)} e^{-i\pi\theta_\Sigma(y_j)}. \tag{2.31}$$

It might be asked at this point what the relation is of the parameters in Eqs. (2.23) and (2.24) and the coupling constant g^2. If this 2D model could be derived from the full 4D theory, a flux-quantization condition might lead to the identification of the vortex charge κ as $\kappa = \pi/g$. In addition, the density ξ would not be an independent parameter but would also be determined by g as $\xi = a^{-2}f(g)$. The inverse temperature β introduced in Eq. (2.23)

is somewhat extraneous since β and κ^2 appear only in the combination $\beta\kappa^2$. Of course, it is not clear that the vortex sector of the 2D effective theory (2.1) is really as simple as this model.

The following heuristic argument is used to suggest that this system has two phases. The energy of an isolated vortex is

$$E=\frac{1}{2}\frac{\kappa^2}{2\pi}\ln\frac{L}{r_0},$$

where L is the linear size of the system. The phase space available to a vortex is the number of lattice sites $(L/a)^2$ so the entropy is

$$S=\ln(L/a)^2.$$

The free energy $F=E-(1/\beta)S$ is

$$F=\left(\frac{\kappa^2}{4\pi}-\frac{2}{\beta}\right)\ln\frac{L}{a}$$

apart from finite terms independent of L. For large β the free energy is positive and divergent as $L\to\infty$ so vortices do not occur as free vortices but bound in $v\bar{v}$ pairs. For small β the free energy is negative and free vortices are preferred. The critical value of β, which is identified as the ionization point, is[28]

$$\beta_c=8\pi/\kappa^2.$$

A more accurate estimate of β_c, which takes into account interactions of $v\bar{v}$ molecules, is[29]

$$\beta_c=\frac{8\pi}{\kappa^2}(1+1.3\pi\xi a^2).$$

The behavior of the expectation value of the loop integral W is now studied for the two phases of the system using natural approximations.

The large-β phase, which $\beta>\beta_c$, is studied in the following approximation: vortices and antivortices are bound in $v\bar{v}$ pairs, and interactions between vortices in different pairs are ignored. This approximation is sensible if the density ξ is small. Then the partition function becomes

$$Z=\sum_n\frac{\xi^{2n}}{(n!)^2}n!\prod_{i=1}^{n}\int d^2x_i d^2y_i e^{\beta\kappa^2 G(x_i-y_i)}, \quad (2.32)$$

where the extra factor $n!$ is the number of pairings of v and $\bar{v}$. Thus

$$Z=\exp\left(\xi^2\int d^2x\,d^2y\,e^{\beta\kappa^2 G(x-y)}\right). \quad (2.33)$$

Use of the approximate form of G in Eq. (2.28) gives

$$Z=\exp\left[\xi^2\Omega\pi r_0^2\left(\frac{\beta\kappa^2}{4\pi}-1\right)^{-1}\right], \quad (2.34)$$

where again $\Omega=\int d^2x$. The specific heat C is

$$C=\xi^2 2\pi r_0^2\left(\frac{\beta\kappa^2}{4\pi}\right)^2\left(\frac{\beta\kappa^2}{4\pi}-1\right)^{-3}. \quad (2.35)$$

The increase of C as β decreases reflects the increase in entropy from fluctuations of the separation lengths of the pairs. The singularity of C at $\beta=4\pi/\kappa^2$ is not real because that value of β is less than β_c, so there the system is in the ionized phase in which the approximations leading to (2.35) are invalid. C is nonsingular at the transition point $\beta_c\simeq 8\pi/\kappa^2$.

The loop integral W, in the approximation expressed by Eq. (2.32), is given by

$$W=\exp\left[-2\xi^2\int d^2x\,d^2y\,e^{\beta\kappa^2 G(x-y)}F_\Sigma(x,y)\right], \quad (2.36)$$

where $F_\Sigma(x,y)$ was defined earlier [see Eq. (2.19)]. The integral can be rewritten by change of variables of integration from x, y to z, ρ, where

$$x=z+\tfrac{1}{2}\rho,\qquad y=z-\tfrac{1}{2}\rho,\qquad \rho=(\rho\cos\theta,\rho\sin\theta), \quad (2.37)$$

as

$$W=\exp\left[-4\pi\xi^2\int_{r_0}^{\infty}\rho\,d\rho\,e^{\beta\kappa^2 G(\rho)}\times\int d^2z\int\frac{d\theta}{2\pi}F_\Sigma(x,y)\right]. \quad (2.38)$$

The integral over z, θ is the same as in Eq. (2.19) and can be evaluated in the limit $L(\partial\Sigma)\to\infty$ as in Eq. (2.20) with the result

$$W\sim\exp\left[-L(\partial\Sigma)4\xi^2 r_0^3\left(\frac{\beta\kappa^2}{4\pi}-\frac{3}{2}\right)^{-1}\right]. \quad (2.39)$$

Thus W obeys a perimeter law in the large-β phase, as expected for a phase in which $v\bar{v}$ bound pairs dominate the partition sum. The difference between this model and the perfect gas of $v\bar{v}$ pairs considered earlier is only that the separation of members of a pair is not fixed at a single value l but has the Boltzmann probability distribution $\exp(-\beta\kappa^2\ln l)$.

In the small-β phase, $\beta<\beta_c$, free vortices occur. It might be expected that W would have an area-law behavior as for a perfect gas of vortices. However, the vortices cannot be treated as a perfect gas because of their long-range Coulomb interaction. This interaction produces a Debye-screening effect. The Debye effect can be taken into account with a technique used by Polyakov,[10] namely by rewriting the theory in the form of a sine-Gordon field theory.[30]

The sine-Gordon form of the theory is derived from Eq. (2.30) for the partition function by use of the following identity of Gaussian functional integration:

$$\int d\varphi(x)\exp\left[-\tfrac{1}{2}c\int d^2x(\nabla\varphi)^2\right]\exp\left[i\int d^2x\,j(x)\varphi(x)\right]=\mathrm{Det}^{-1/2}(-c\nabla^2)\exp\left[-\frac{1}{2c}\int d^2x\,d^2y\,j(x)G(x-y)j(y)\right], \tag{2.40}$$

where $\varphi(x)$ is a lattice scalar field. If this identity is applied to Eq. (2.30) with

$$j(x)=\sum_\lambda e_\lambda\delta^2(x-x_\lambda), \qquad c=\frac{1}{\beta\kappa^2}, \tag{2.41}$$

then Z can be written

$$Z=e^{2\xi\Omega}\,\mathrm{Det}^{1/2}\left(-\frac{1}{\beta\kappa^2}\nabla^2\right)\int d\varphi(x)e^{-S(\varphi)}, \tag{2.42}$$

where

$$S(\varphi)=\int d^2x\left[\frac{1}{2\beta\kappa^2}(\nabla\varphi)^2+2\xi(1-\cos\varphi)\right]. \tag{2.43}$$

Similarly, the sine-Gordon expression for W is

$$W=\int d\varphi(x)e^{-S_\Sigma(\varphi)}\Big/\int d\varphi(x)e^{-S(\varphi)}, \tag{2.44}$$

where

$$S_\Sigma(\varphi)=\int d^2x\left\{\frac{1}{2\beta\kappa^2}(\nabla\varphi)^2+2\xi[1-\cos(\varphi+\pi\theta_\Sigma)]\right\} \tag{2.45}$$

and the function $\theta_\Sigma(x)$ is defined in Eq. (2.6).

For β small the gradient term in the action $S(\varphi)$ forces φ to be smooth. It makes sense to expand the cosine in Eq. (2.43) and keep terms only of order φ^2. In that Gaussian approximation the partition function becomes

$$Z=e^{2\xi\Omega}\,\mathrm{Det}^{1/2}\left(-\frac{1}{\beta\kappa^2}\nabla^2\right)\mathrm{Det}^{-1/2}\left(-\frac{1}{\beta\kappa^2}\nabla^2+2\xi\right). \tag{2.46}$$

The free energy is

$$\beta F=-\frac{1}{\Omega}\ln Z=-2\xi+\tfrac{1}{2}\int\frac{d^2p}{(2\pi)^2}\ln\frac{p^2+2\xi\beta\kappa^2}{p^2} \tag{2.47}$$

and the specific heat C is

$$C=\xi\frac{\beta\kappa^2}{4\pi}. \tag{2.48}$$

The linear increase of C with β implies a constant rate of increase of entropy as β decreases.

The loop integral W is evaluated in a similar way by making a Gaussian approximation of the functional integral in Eq. (2.44). Here it is necessary to expand φ around the field φ_C that minimizes $S_\Sigma(\varphi)$; that is, write $\varphi=\varphi_C+\chi$ and expand $S_\Sigma(\varphi)$ up to terms quadratic in χ. In that approximation the formula for W is

$$W=e^{-S_\Sigma(\varphi_C)}\tilde{W}, \tag{2.49}$$

where

$$\tilde{W}=\mathrm{Det}^{+1/2}\left(-\frac{1}{\beta\kappa^2}\nabla^2+2\xi\right)\times\mathrm{Det}^{-1/2}\left[-\frac{1}{\beta\kappa^2}\nabla^2+2\xi\cos(\varphi_C+\pi\theta_\Sigma)\right]. \tag{2.50}$$

The field $\varphi_C(x)$ that minimizes $S_\Sigma(\varphi)$ is not known to me. However, a qualitative picture of φ_C can be obtained by the following simply considerations. The gradient term in $S_\Sigma(\varphi)$ is small if φ is a slowly varying field; the other term is small if $\varphi\simeq-\pi\theta_\Sigma$. If Σ is a large region then the balance between these two terms will produce a field φ_C equal to $-\pi$ inside Σ, equal to 0 outside Σ, and with a smooth transition from $-\pi$ to 0 over a region of thickness ϵ around the boundary $\partial\Sigma$ of Σ. The action of that field, as a function of ϵ, is approximately

$$S_\Sigma\simeq\frac{1}{2\beta\kappa^2}L(\partial\Sigma)\epsilon\left(\frac{\pi}{\epsilon}\right)^2+2\xi L(\partial\Sigma)\epsilon. \tag{2.51}$$

This is minimized by the choice

$$\epsilon=\frac{\pi}{2}\lambda_D, \tag{2.52}$$

where

$$\lambda_D=(\xi\beta\kappa^2)^{-1/2}, \tag{2.53}$$

and for that value of ϵ the action is

$$S_\Sigma(\varphi_C)=L(\partial\Sigma)2\pi\xi\lambda_D. \tag{2.54}$$

The approximations leading to this estimate are valid if the region Σ is large compared to the length λ_D. The length λ_D might be called the Debye length of the 2D plasma of vortices. It is the characteristic limiting distance over which fluctuations of the densities of vortices and antivortices can extend.

To complete the estimate of W it is necessary to consider the factor $\tilde{W}$. The logarithm of $\tilde{W}$ can be written

$$\ln\tilde{W}=-\tfrac{1}{2}\,\mathrm{Tr}\ln\left[-\frac{1}{\beta\kappa^2}\nabla^2+2\xi\cos(\varphi_C+\pi\theta_\Sigma)\right]+\tfrac{1}{2}\,\mathrm{Tr}\ln\left(-\frac{1}{\beta\kappa^2}\nabla^2+2\xi\right)$$

$$=-\tfrac{1}{2}\int dE\,\ln E[\rho(E)-\rho_0(E)], \qquad (2.55)$$

where $\rho(E)$ and $\rho_0(E)$ are the densities of states of the differential operators. A reasonable qualitative estimate of $\tilde{W}$ can be obtained by using the Thomas-Fermi approximation for $\rho(E)$ (Ref. 31):

$$\rho(E)=\int d^2x\int\frac{d^2p}{(2\pi)^2}\delta\left(E-\frac{p^2}{\beta\kappa^2}-2\xi\cos[\varphi_C(x)+\pi\theta_\Sigma(x)]\right). \qquad (2.56)$$

This gives

$$\ln\tilde{W}=\tfrac{1}{2}\int d^2x\int\frac{d^2p}{(2\pi)^2}\ln\frac{p^2+2\xi\beta\kappa^2}{p^2+2\xi\beta\kappa^2\cos(\varphi_C+\pi\theta_\Sigma)}. \qquad (2.57)$$

The function $\cos(\varphi_C+\pi\theta_\Sigma)$ is approximately equal to 1 unless x is in the region within the distance ϵ of $\partial\Sigma$; in that region the cosine is <1 and might be approximated by zero. Thus, qualitatively $\ln\tilde{W}$ is

$$\ln\tilde{W}=\tfrac{1}{2}L(\partial\Sigma)\epsilon\int\frac{d^2p}{(2\pi)^2}\ln\left(1+\frac{2\xi\beta\kappa^2}{p^2}\right). \qquad (2.58)$$

The momentum variable p is restricted by $|p|\leq\pi/a$ because of the lattice (which has been suppressed up to now), so

$$\ln\tilde{W}\simeq L(\partial\Sigma)\tfrac{1}{8}\xi\lambda_D\beta\kappa^2[1-\ln(\xi a^2\beta\kappa^2)+\cdots], \qquad (2.59)$$

where the terms that have been dropped are of higher order in $\beta\kappa^2$, which is assumed small. Comparison of $\ln\tilde{W}$ to $S_\Sigma(\varphi_C)$ shows that $\ln\tilde{W}$ is smaller by a factor of order $\beta\kappa^2\ln\beta\kappa^2$. Thus to lowest order in $\beta\kappa^2$ the factor $\tilde{W}$ can be ignored, and W is, for $L(\partial\Sigma)\gg\lambda_D$,

$$W\sim e^{-L(\partial\Sigma)/2\pi\xi\lambda_D}. \qquad (2.60)$$

Because of the Debye screening W obeys a perimeter law in the small-β phase of the 2D Coulomb gas.

In the limit $\beta\to 0$, W must approach the area law Eq. (2.14) associated with a perfect gas of vortices. In the limit $\beta\to 0$ the Debye length $\lambda_D\to\infty$. The perimeter law (2.60) holds for a loop $\partial\Sigma$ large compared to λ_D. More generally, it is expected that

$$W\simeq\begin{cases}e^{-\xi A(\Sigma)} & \text{for } l(\Sigma)\lesssim\lambda_D,\\ e^{-\xi L(\partial\Sigma)/\lambda_D} & \text{for } l(\Sigma)\gg\lambda_D\end{cases} \qquad (2.61)$$

apart from constants, where $l(\Sigma)$ is the linear size of Σ. In the limit $\beta\to 0$ the region over which the area law holds becomes infinite as $\lambda_D\to\infty$, and in this way the area law of the perfect gas of vortices is recovered.

Debye screening reduces the topological disorder created by the ionized vortices. Fluctuations of v and $\bar{v}$ densities are limited to regions of size $\lesssim\lambda_D$. Thus for Σ large compared to λ_D the vortices that affect W are those that lie within a distance λ_D of $\partial\Sigma$. This leads to the perimeter law Eq. (2.60).

To summarize, in statistical-mechanics models of the 2D vortex contribution to the loop integral W it is found that an area-law dependence $W\sim\exp(-A(\Sigma))$ is associated with existence of free vortices, and perimeter law $W\sim\exp(-L(\partial\Sigma))$ with dominance of correlated $v\bar{v}$ pairs. The correlation needed to give perimeter-law behavior need not be as strong as requiring every vortex to be bound to an antivortex; Debye-screening correlations are sufficient. The distinction between area law and perimeter law measures topological disorder: a $v\bar{v}$ pair contributes weakly to W because the long-range vortex and antivortex fields *cancel*, leaving just a short-range field and no topological disorder.

III. LONG-RANGE VACUUM GAUGE FIELDS

In Sec. II the expectation value of the Wilson loop integral W is examined for several *models* of the vortex sector of the effective two-dimensional theory with probability distribution $P_2(\varphi_a^\alpha)$ defined in Eq. (2.1). The purpose of the present section is to consider the effective *three*-dimensional theory with probability distribution $P_3(\varphi_a^i)$ derived from the full 4D theory by an equation analogous to (2.1). To be specific, the effect of long-range magnetic monopole fields in the 3D theory, and their role in creating vortices in the 2D theory, will be discussed. In addition, a brief description of operators dual to the loop operator W, such as the dual operator defined by 't Hooft,[21] will be presented in the context of the effective 3D theory.

The reduced probability distribution $P_3(A_a^i)$ of the 3D theory is the distribution of the field configurations $A_a^i(\vec{x})$ $(i=1,2,3)$ on the 3D subspace with $x_4=0$. It is defined by the path integral [compare Eq. (2.1)]

$$P_3(A_a^i)=\int dA_a'^\mu(x)e^{-S(A_a')}\prod_{a,i=1}^{3}\prod_{\vec{x}}\delta[A_a^i(\vec{x})-A_a'^i(\vec{x},0)], \qquad (3.1)$$

where again the gauge-fixing terms and Faddeev-Popov determinant have been suppressed.

The partition function Z can be written in terms of $P_3(A_a^i)$ as

$$Z=\int dA_a^i(\vec{x})P_3(A_a^i)\,. \tag{3.2}$$

The expectation value of the loop integral W, for a surface Σ in the $x_4=0$ subspace, can be written

$$W=Z^{-1}\int dA_a^i(\vec{x})P_3(A_a^i)\tfrac{1}{2}\,\mathrm{Tr}P\exp\left(ig\oint_{\partial\Sigma}ds_i A_a^i t_a\right). \tag{3.3}$$

Thus Z and W are determined by $P_3(A_a^i)$. Also, the effective 2D theory can be obtained from $P_3(A_a^i)$: the 2D probability distribution $P_2(\varphi_a^\alpha)$ is

$$P_2(\varphi_a^\alpha)=\int dA_a^i(\vec{x})P_3(A_a^i)\prod_{a,\alpha}\prod_{x_\perp}\delta[\varphi_a^\alpha(x_\perp)-A_a^\alpha(x_\perp,0)], \tag{3.4}$$

where $x_\perp$ is the position in the (12) plane and α takes the values 1, 2.

The probability distribution $P_3(A_a^i)$ can be interpreted as the square of the Schrödinger wave functional of the vacuum state

$$P_3(A_a^i)=\Psi^2(A_a^i),$$

where

$$\Psi(A_a^i)=\langle A_a^i|0\rangle$$

with $|0\rangle$ the vacuum state and $|A_a^i\rangle$ the eigenstate of the gauge-field operator with eigenvalue $A_a^i(\vec{x})$. The identification $P_3=\Psi^2$ can be derived from the path-integral formula (3.1) in the manner described in Ref. 14. The distribution of vacuum fluctuations of the gauge fields $P_3(A_a^i)$ will be referred to as the vacuum functional.

The full 4D theory is needed to understand the vacuum functional $P_3(A_a^i)$; and $P_3(A_a^i)$ in turn determines the distribution of 2D fields $P_2(\varphi_a^\alpha)$ from which W is found. By this successive reduction of dimensions, the topological configurations of each dimension derive from those of one higher dimension. In this section, 3D monopole fields are considered as the origin of 2D vortexlike configurations. Of course, there are other 3D fields that would produce vortices in the 2D subspace $x_3=0$; for example, 3D vortex loops. However, monopole fields may be the most economical way to create the topological disorder in the 2D theory that is associated with the area law $W\sim\exp(-A(\Sigma))$.

The study of the 3D monopole sector to be described follows the earlier study of the 2D vortex sector. That is, the integral $\int dA_a^i(\vec{x})$ is replaced by integrals over just monopole coordinates, and several simple models of the distribution of monopoles are used to study the influence of monopoles on W.

The monopole field $\vec{A}_m(\vec{x})$ in an Abelian gauge theory is the Dirac monopole field

$$A_m^i(\vec{x})=-\frac{1}{2g}\epsilon_{3ij}\frac{x_j}{r(r+x_3)}\,, \tag{3.5}$$

where $r=|\vec{x}|$ and the monopole is centered at $\vec{x}=0$. This field is singular along the negative x_3 axis, which is the position of the Dirac string. It obeys the Dirac magnetic-charge quantization condition

$$\int_{S_\epsilon}d\vec{\sigma}\cdot(\nabla\times\vec{A}_m)=\frac{2\pi}{g}\,, \tag{3.6}$$

where S_ϵ is a spherical surface surrounding the origin *minus* an infinitesimal cap of radius ϵ around the negative x_3 axis. Equivalently, the flux of $\nabla\times\vec{A}_m$ carried into the origin by the Dirac string is $2\pi/g$.

The string singularity in Eq. (3.5) is along the x_3 axis. Alternatively, it can be directed along any other curve from $\vec{x}=0$ to infinity. Since monopole fields with different string positions are gauge equivalent, there is no loss of generality in letting the string be along the negative x_3 axis. Polyakov has shown[16] that monopole fields with quantization condition (3.6) arise naturally in Abelian lattice gauge theories. The string is unobservable because in the lattice theory A^i appears only in the operators $\exp(ig\int_\square ds\cdot A)$ where $\square$ represents an elementary plaquette of the lattice; the contribution of the string part is $\exp(i2\pi)=1$, so is irrelevant.

A 3D monopole creates a 2D vortex. The field $A_m^i(\vec{x})$ on the $x_3=0$ subspace is precisely the vortex configuration $\varphi_v^\alpha(x_\perp)$ in Eq. (2.4):

$$A_m^\alpha(x_\perp,0)=-\frac{1}{2g}\epsilon_{\alpha\beta}\frac{x_\beta}{x_\perp^2}\,, \tag{3.7}$$

where $\alpha=1,2$. In particular, the Dirac magnetic-charge quantization condition leads to the vortex flux quantization (2.5) needed to make the loop integral $W=-1$.

The monopole $A_m^i(\vec{x})$ is centered at $\vec{x}=0$. More generally, if the monopole center is $\vec{c}$, i.e., the field is $A_m^i(\vec{x}-\vec{c})$, the value of the field on the $x_3=0$ subspace is

$$A_m^\alpha(x_\perp-c_\perp;-c_3)=-\frac{1}{2g}\epsilon_{\alpha\beta}\frac{x_\beta-c_\beta}{(x_\perp-c_\perp)^2-c_3[(x_\perp-c_\perp)^2+c_3^2]^{1/2}+c_3^2}\,. \tag{3.8}$$

This field has some of the same character as the point vortex (3.7) with its center shifted to $c_\perp$; but it has also the qualitative difference that it depends on the additional parameter c_3. In the 2D field, c_3 is a *scale* parameter. That is, A_m^α is asymptotically of the form (3.7) and c_3 sets the scale of the asymptotic regime:

$$A_m^\alpha \sim \varphi_v^\alpha(x_\perp - c_\perp) \quad \text{for } |x_\perp - c_\perp| \gg c_3 .$$

It follows that the loop integral $g\oint_{\partial\Sigma} ds_\alpha A_m^\alpha$ approaches π only for $\partial\Sigma$ large compared to c_3.

The loop integral for a monopole centered at $\vec{x}$ will be denoted $\Phi_\Sigma(\vec{x})$:

$$\Phi_\Sigma(\vec{x}) = g\oint_{\partial\Sigma} d\vec{s}\cdot\vec{A}_m(\vec{s}-\vec{x}) = g\int_\Sigma d\vec{\sigma}\cdot\vec{B}(\vec{\sigma}-\vec{x}) \tag{3.9}$$

and again the loop $\partial\Sigma$ is taken to lie in the (12) plane. Since $\Phi_\Sigma(\vec{x})$ is the magnetic flux through Σ due to a point monopole at $\vec{x}$, it can be written

$$\Phi_\Sigma(\vec{x}) = \tfrac{1}{2}\omega_\Sigma(\vec{x}), \tag{3.10}$$

where $\omega_\Sigma(\vec{x})$ is defined as the solid angle at $\vec{x}$ subtended by the surface Σ. In Eq. (3.10) the magnetic flux of the Dirac string has been dropped since it is $2\pi\theta(-x_3)$ and thus contributes nothing to $e^{i\Phi_\Sigma}$, which is what appears in W. The solid angle $\omega_\Sigma(\vec{x})$ cannot be evaluated analytically for arbitrary Σ and $\vec{x}$. Some qualitative features of $\omega_\Sigma(\vec{x})$ are as follows. For $\vec{x}$ near Σ, i.e., $x_\perp\in\Sigma$ and $x_3 \ll l(\Sigma)$ ($\equiv$ size of Σ), the solid angle $\omega_\Sigma(\vec{x}) \simeq 2\pi$; then $e^{i\Phi_\Sigma}\simeq -1$ as for a 2D *point* vortex. For $\vec{x}$ far from Σ, i.e., $|\vec{x}| \gg l(\Sigma)$, the solid angle $\omega_\Sigma(\vec{x})$ is asymptotically

$$\omega_\Sigma(\vec{x}) \sim \frac{A(\Sigma)}{r^2}\cos\theta, \tag{3.11}$$

where (r,θ,φ) are the polar coordinates of $\vec{x}$. Finally, if Σ is a disk of radius R centered at the origin of the (1, 2) plane and $\vec{x}$ is on the x_3 axis,

$$\omega_\Sigma(0,0,x_3) = 2\pi\left(1 - \frac{x_3}{(R^2+x_3^2)^{1/2}}\right). \tag{3.12}$$

Thus ω_Σ varies from 0 for $x_3 \gg R$ to 2π for $x_3 \ll R$; this shows the role of x_3 as a scale parameter.

In the Abelian theory an antimonopole field is defined as $-A_m^i(\vec{x})$. Also, multimonopole fields are obtained by superposition; the general field is

$$\bar{A}^i(\vec{x}) = \sum_{i=1}^{n} A_m^i(\vec{x}-\vec{x}_i) - \sum_{j=1}^{n'} A_m^i(\vec{x}-\vec{y}_j), \tag{3.13}$$

where $\vec{x}_i$ and $\vec{y}_j$ are the positions of n monopoles and n' antimonopoles. The loop integral of this field is

$$\exp\left(ig\oint_{\partial\Sigma} ds_i \bar{A}^i\right) = \prod_{i=1}^{n} e^{i\Phi_\Sigma(x_i)} \prod_{j=1}^{n'} e^{-i\Phi_\Sigma(y_j)} . \tag{3.14}$$

There are monopole field configurations in non-Abelian gauge theories. First, the Dirac monopole field $A_m^i(\vec{x})$ can be embedded in the SU(2) gauge theory by defining

$$A_{ma}^i(\vec{x}) = 2n_a A_m^i(\vec{x}), \tag{3.15}$$

where n_a is a constant unit vector, $n_a^2 = 1$. This field is singular on the Dirac string of $A_m^i(\vec{x})$. On the other hand, there exists an SU(2) gauge transformation that eliminates the Dirac string and yields a nonsingular field which is just the Wu-Yang monopole configuration[4,32]

$$A_{ma}^i(\vec{x}) = \frac{1}{g}\epsilon_{aij}\frac{x^j}{\vec{x}^2} . \tag{3.16}$$

These configurations create *point* vortices when restricted to the 2D subspace $x_3 = 0$ just as in the Abelian case; and the resulting non-Abelian vortices satisfy precisely the flux-quantization condition (2.5) needed to make

$$\tfrac{1}{2}\mathrm{Tr}P\exp\left(ig\oint_{\partial\Sigma} d\vec{s}\cdot\vec{A}_a t_a\right) = -1 .$$

Also, an SU(2) monopole field $A_{ma}^i(\vec{x}-\vec{c})$ centered at $\vec{c}$ creates an SU(2) vortex with *scale parameter* c_3 in the 2D subspace, as described above.

The loop integrals of the SU(2) monopole fields (3.15) and (3.16) are equal because the fields are gauge equivalent, and given by

$$\tfrac{1}{2}\mathrm{Tr}P\exp\left[ig\oint_{\partial\Sigma} ds^i A_{ma}^i(\vec{s}-\vec{x})t_a\right] = \cos\Phi_\Sigma(\vec{x}), \tag{3.17}$$

where $\vec{x}$ is the monopole position and $\Phi_\Sigma(\vec{x})$ is defined in Eq. (3.9). For $|\vec{x}| \ll l(\Sigma)$, where $l(\Sigma)$ is the linear size of Σ, $\cos\Phi_\Sigma \sim -1$; thus the Wu-Yang monopole field obeys the charge-quantization condition needed to have a large influence on W. Note, however, that for adjoint-representation generators T_a of SU(2) the loop integral is

$$\tfrac{1}{3}\mathrm{Tr}P\exp\left(ig\oint_{\partial\Sigma} ds^i A_a^i T_a\right) = \cos 2\Phi_\Sigma ,$$

which approaches 1 as $l(\Sigma)\to\infty$; the monopole field does not have a large effect on the loop operator in the adjoint representation.

Presumably there exist multimonopole configurations in the non-Abelian gauge theory, analogs of the Abelian field in Eq. (3.13), with additional degrees of freedom corresponding to relative global SU(2) orientations. It is not known to me how to construct a general multimonopole field. In the simple models of the monopole sector considered below, the Abelian formula Eq. (3.14) is used for the loop integral of a multimonopole field[33] and the degrees of freedom associated with a monopole are just its position. Such a model

is probably a poor excuse for a description of multimonopole fields in a non-Abelian gauge theory. Still, it may be that the statistical mechanics of a gas of Abelian monopoles is a first step toward a theory of the influence of long-range gauge fields.

Statistical-mechanics models of the monopole sector of the theory are obtained by replacing the functional integral $\int dA_a^i(\vec{x})$ by integrals just over monopole coordinates. The partition function Z becomes

$$Z=\sum_{n,n'}\frac{1}{n!\,n'!}\int\prod_{i,j=1}^{n,n'}d^3x_i d^3y_j P(x,y) \tag{3.18}$$

and the loop integral W

$$W=Z^{-1}\sum_{n,n'}\frac{1}{n!\,n'!}\int\prod_{i,j=1}^{n,n'}d^3x_i d^3y_j P(x,y)\times e^{i\Phi_\Sigma(\vec{x}_i)}e^{-i\Phi_\Sigma(\vec{y}_j)}. \tag{3.19}$$

Here $\vec{x}_i$ and $\vec{y}_j$ are monopole and antimonopole positions, the factor $1/n!\,n'!$ is the Boltzmann counting factor needed because fields that differ only by interchange of $\vec{x}_i$'s or $\vec{y}_j$'s are identical, and $P(x,y)$ is the probability distribution of monopoles. The integrand of W is the loop integral of a multimonopole field, Eq. (3.14). The function $P(x,y)$, which might be derived in principle as the probability distribution of multimonopole fields in the full functional integral, will here be chosen *ad hoc* for a few simple models.

Before discussing these models it is useful to introduce another quantity M that is similar to the loop integral W. Let M be defined by

$$M=Z^{-1}\sum_{n,n'}\frac{1}{n!\,n'!}\int\prod_{i,j}d^3x_i d^3y_j P(x,y)\times e^{i\pi\theta_V(\vec{x}_i)}e^{-i\pi\theta_V(\vec{y}_j)}, \tag{3.20}$$

where V is a finite region of the 3D space and the function $\theta_V(\vec{x})$ is defined as

$$\theta_V(\vec{x})=\begin{cases}1 & \text{if } \vec{x}\in V,\\ 0 & \text{if } \vec{x}\notin V.\end{cases} \tag{3.21}$$

The quantity M is interesting for various reasons. First, M is more specifically related to the monopole fields than W: M depends on $\pi\theta_V$ which is π times the number of monopoles in V, while W depends on Φ_Σ, the flux of $\nabla\times\vec{A}$ through the 2D surface Σ. Second, M is the generalization to the 3D monopole models of W in the 2D vortex models: compare their definitions Eqs. (3.20) and (2.11). Third, the dependences of M on V and W on Σ are related, and either can be used as a measure of topological disorder. Finally, in the discussion of the XY model in Sec. V, analogs of both W and M will be defined.

The quantity M could be defined in the full field theory as the expectation value of $\exp(\frac{1}{2}ig\int'_{\partial V}d\vec{\sigma}\cdot(\nabla\times\vec{A}))$ where the surface integral $\int'_{\partial V}$ is over the boundary surface ∂V of V minus an infinitesimal surface around each Dirac string. M measures the magnetic flux through the closed surface ∂V, whereas W measures the flux through the open surface Σ.

The discussion of the influence of monopoles on W will begin with a perfect gas of monopoles. The probability distribution is taken to be

$$P(x,y)=\rho^{n+n'}, \tag{3.22}$$

where ρ, which has units (length)$^{-3}$, is the density. Since $P(x,y)$ is independent of the positions x_i and y_j, the monopoles are uncorrelated. The partition function is

$$Z=e^{2\rho\Omega}, \tag{3.23}$$

where now $\Omega=\int d^3x$ is the volume of three-space, kept finite for the moment.

The loop integral W is

$$W=\exp\left[-2\rho\int d^3x[1-\cos\Phi_\Sigma(\vec{x})]\right]. \tag{3.24}$$

For simplicity, the surface Σ will be taken to be a disk of radius R centered at the origin in the (12) plane. Even for this case the integral in Eq. (3.24) cannot be evaluated analytically. However, by dimensional analysis the integral is proportional to R^3. Thus

$$W=\exp(-R^3 2\rho K), \tag{3.25}$$

where K is a numerical factor of order 1 defined by

$$K=\int d^3x[1-\cos\tfrac{1}{2}\omega_1(\vec{x})], \tag{3.26}$$

where $\omega_1(\vec{x})$ is the solid angle at $\vec{x}$ subtended by a disk of radius 1 at the origin; $\omega_1(\vec{x})$ decreases sufficiently rapidly as $|\vec{x}|\to\infty$ so that K is finite.

Equation (3.25) is an interesting result. The loop integral decreases exponentially as R^3, i.e., more rapidly than an area law. The topological disorder created by a perfect gas of monopoles is even greater than what is needed to satisfy the conconfinement criterion $W\sim\exp(-R^2)$.

The origin of the R^3 dependence in Eq. (3.25) can be explained as follows. The monopoles create 2D vortices with all values of scale parameter x_3. The positions and scales of the vortices are uncorrelated. Roughly, vortices with scale less than R contribute to the loop integral as point vortices, and those with scale greater than R do not affect the loop integral. Thus, $W\simeq\exp(-\xi R^2)$, where ξ is the density of vortices with scale less

than R, $\xi \simeq \rho R$; so $W \sim \exp(-\rho R^3)$.

Next the quantity M is considered. For the perfect gas it is

$$M = \exp(-4\rho V), \tag{3.27}$$

where $V = \int d^3x\, \theta_V(\vec{x})$ is the volume of the region V. This dependence of M on the volume V is like the dependence of W on the area $A(\Sigma)$ in the case of a perfect gas of 2D point vortices [Eq. (2.14)].

If the region Σ is a disk and V a sphere both of radius R, then both W and M decrease as $\exp(-R^3)$. This follows from the similarity of the two functions $\Phi_\Sigma(\vec{x})$ and $\pi\theta_V(\vec{x})$. To be specific, $\Phi_\Sigma(\vec{x})$ is approximately equal to π if $|\vec{x}| \lesssim R$, and decreases rapidly to 0 as $|\vec{x}|/R \to \infty$; the function $\pi\theta_V(\vec{x})$ takes those values exactly. This similarity justifies the remark that either W or M can be used to measure the topological disorder of these models.

The next model that will be considered is a perfect gas of monopole-antimonopole bound pairs, with the separation of m and $\bar{m}$ in a pair fixed at l. The degrees of freedom of each pair are the position $\vec{z}$ of the center of mass, and the angular orientation described by polar angles (θ, φ). The partition function of this system is

$$Z = \sum_{n=0}^{\infty} \frac{\rho'^n}{n!} \prod_{i=1}^{n} \int d^3z_i \int \frac{d\Omega_i}{4\pi}, \tag{3.28}$$

where $d\Omega_i = \sin\theta_i d\theta_i d\varphi_i$ and ρ' is the density; thus

$$Z = e^{\rho' \Omega}. \tag{3.29}$$

The loop integral in this model is

$$W = \exp\left\{-\rho' \int d^3z \int \frac{d\Omega}{4\pi}[1 - \cos(\Phi_\Sigma(\vec{x}) - \Phi_\Sigma(\vec{y}))]\right\}, \tag{3.30}$$

where $\vec{x}$ and $\vec{y}$ are the m and $\bar{m}$ positions associated with $m\bar{m}$ coordinates (z, θ, φ):

$$\begin{aligned} \vec{x} &= \vec{z} + \tfrac{1}{2}\vec{\rho}, \\ \vec{y} &= \vec{z} - \tfrac{1}{2}\vec{\rho} \end{aligned} \tag{3.31}$$

and the polar coordinates of $\vec{\rho}$ are (l, θ, φ). When the region Σ is large compared to the separation l, the integrand in Eq. (3.30) is small unless *one* of $\vec{x}$ and $\vec{y}$ lies near Σ; this only occurs if $\vec{z}$ lies in a tube of diameter l around the boundary $\partial\Sigma$ of Σ. Thus, for $R \gg \Sigma$ the quantity W is of the form

$$W \sim \exp[-L(\partial\Sigma)\rho'\pi l^2 K'], \tag{3.32}$$

where K' is a numerical factor of order 1. Thus W has perimeter-law behavior in this model. The origin of the perimeter dependence of W is the strict correlation of vortices in the 2D slice of the theory: vortices are combined in $v\bar{v}$ pairs with positions x, y and scales x_3, y_3 such that $[(x-y)^2 + (x_3 - y_3)^2]^{1/2} = l$. Lack of topological disorder yields the perimeter law (3.32).

In this model the quantity M is

$$M = \exp\left\{-\rho' \int d^3z \int \frac{d\Omega}{4\pi}\{1 - \cos\pi[\theta_V(\vec{x}) - \theta_V(\vec{y})]\}\right\}. \tag{3.33}$$

The integrand in Eq. (3.33) vanishes unless exactly one of $\vec{x}$ and $\vec{y}$ lies in the volume V. Thus M is again analogous to the quantity W in the 2D model of a perfect gas of $v\bar{v}$ pairs [Eq. (2.17)]. It can be shown that in the limit in which the size of V is large compared to l, M approaches

$$M \sim \exp(-S\rho' \tfrac{1}{8} l), \tag{3.34}$$

where S is the area of the surface ∂V that is the boundary of V. The strict correlation of m and $\bar{m}$ positions reduces the dependence of M on the region V from the volume dependence Eq. (3.27) of a perfect gas to surface-area dependence Eq. (3.34).

To summarize the models considered above, it can be said that the dependence of either W on Σ or M on V distinguishes between topological order and disorder. The effect of the long-range monopole field on W and M leads to dependence of these quantities on the volume R^3 or V. In contrast, the short-range $m\bar{m}$ dipole fields affect W and M only if the dipole lies near the boundary of Σ or V, which leads to dependence of W on $L(\partial\Sigma)$ and M on S.

The final model that will be considered is a 3D Coulomb gas of monopoles and antimonopoles. There are probably strong correlations of the collective coordinates of long-range vacuum fluctuations in confining gauge theories, but the correlation must be less than that of models in which monopoles are bound in $m\bar{m}$ dipole pairs. In the 3D Coulomb gas, the interaction of monopoles produces correlations of monopole positions intermediate between that of the two models considered above.

The probability distribution of a Coulomb gas of monopoles is

$$P(x, y) = \rho^{n+n'} \exp\left(-\beta \tfrac{1}{2} \sum_{\lambda \neq \lambda'} \frac{\kappa^2}{4\pi} \frac{e_\lambda e_{\lambda'}}{|\vec{x}_\lambda - \vec{x}_{\lambda'}|}\right), \tag{3.35}$$

where $(e_\lambda, \vec{x}_\lambda)$ stands for $(+1, \vec{x}_i)$ for $\lambda = i = 1, \ldots, n$, and stands for $(-1, \vec{y}_j)$ for $\lambda = n + j$, $j = 1, \ldots, n'$. The parameter κ should be identified with $2\pi/g$, the Dirac magnetic charge of the monopoles; ρ is the density. The Coulomb interaction (3.35) is attractive (repulsive) for $m\bar{m}$ (mm or $\bar{m}\bar{m}$) combinations.

The Coulomb interaction produces Debye screening in a 3D plasma. The Debye length λ_D is

$$\lambda_D = (\rho\beta\kappa^2)^{-1/2} . \tag{3.36}$$

In analogy with the 2D problem [Eq. (2.31)] it is expected that the correlation of monopole positions responsible for Debye screening reduces the topological disorder in comparison with a perfect gas, and thus reduces the rate of decrease of W or M as R, the linear size of Σ or V, tends to infinity.

The Debye screening can be treated by a method used by Polyakov.[10] The starting point is to transform the system to an equivalent sine-Gordon field theory, in the same way as for the 2D Coulomb gas. That is, the 3D form of the Gaussian functional integral formula Eq. (2.40) is used to derive a formula for the partition function Z in terms of a scalar field $\varphi(\vec{x})$:

$$Z = e^{2\rho\Omega}\,\mathrm{Det}^{1/2}\left(-\frac{1}{\beta\kappa^2}\nabla^2\right)\int d\varphi(\vec{x})e^{-S(\varphi)}, \tag{3.37}$$

where

$$S(\varphi) = \int d^3x\left[\frac{1}{2\beta\kappa^2}(\nabla\varphi)^2 + 2\rho(1-\cos\varphi)\right]. \tag{3.38}$$

Similar derivations lead to expressions for the quantities W and M in the sine-Gordon theory

$$W = \int d\varphi(\vec{x})e^{-S_\Sigma(\varphi)} \Big/ \int d\varphi(\vec{x})e^{-S(\varphi)}, \tag{3.39a}$$

where

$$S_\Sigma(\varphi) = \int d^3x\left\{\frac{1}{2\beta\kappa^2}(\nabla\varphi)^2 + 2\rho[1-\cos(\varphi+\Phi_\Sigma)]\right\}; \tag{3.39b}$$

and

$$M = \int d\varphi(\vec{x})e^{-S_V(\varphi)} \Big/ \int d\varphi(\vec{x})e^{-S(\varphi)}, \tag{3.40a}$$

where

$$S_V(\varphi) = \int d^3x\left\{\frac{1}{2\beta\kappa^2}(\nabla\varphi)^2 + 2\rho[1-\cos(\varphi+\pi\theta_V)]\right\}. \tag{3.40b}$$

The 3D Coulomb gas of monopoles [Eqs. (3.37)–(3.39)] has been described by Polyakov.[10] He derived this system from the (2 + 1)-dimensional Georgi-Glashow model, an SU(2) gauge theory with Higgs fields; the monopole solutions that occur in that model[32] play the role of instantons, and Eq. (3.18) for Z originates as an expansion in instanton coordinates.

Equations (3.39) and (3.40) show that the quantities M and W are similar. The crucial point of similarity is that the functions $\pi\theta_V(\vec{x})$ and $\Phi_\Sigma(\vec{x})$ that define M and W are both discontinuous over a 2D surface, namely, ∂V and Σ. Because of the discontinuity, M and W decrease exponentially as the *area* of ∂V and Σ.

To study the quantity M, the integrals $\int d\varphi(\vec{x})$ will be evaluated in the Gaussian approximation, which is valid for $\beta\kappa^2$ small: Let φ_c be the field that minimizes $S_V(\varphi)$, shift the integration variable from $\varphi(\vec{x})$ to $\chi(\vec{x}) = \varphi(\vec{x}) - \varphi_c(\vec{x})$, and expand S_V up to terms quadratic in χ. The quantity M becomes

$$M = e^{-S_V(\varphi_c)}\tilde{M}, \tag{3.41}$$

where

$$\tilde{M} = \mathrm{Det}^{1/2}\left(-\frac{1}{\beta\kappa^2}\nabla^2 + 2\rho\right) \times \mathrm{Det}^{-1/2}\left(-\frac{1}{\beta\kappa^2}\nabla^2 + 2\rho\cos(\varphi_c+\pi\theta_V)\right). \tag{3.42}$$

A qualitative understanding of the dependence of M on the region V can be obtained by an analysis that parallels that carried out earlier [Eq. (2.49)] for the 2D Coulomb gas. For V large compared to the Debye length λ_D, the field $\varphi_c(\vec{x})$ that minimizes $S_V(\varphi)$ is approximately equal to $-\pi$ for $\vec{x}$ in V, to 0 for $\vec{x}$ not in V, with a transition from $-\pi$ to 0 in a region of thickness $(\pi/2)\lambda_D$ around the boundary surface ∂V of V. The action of $\varphi_c(\vec{x})$ is approximately

$$S_V(\varphi_c) \simeq S2\pi\rho\lambda_D , \tag{3.43}$$

where again S is the surface area of ∂V.

The quantity $\tilde{M}$ can be estimated from the Thomas-Fermi approximation of the density of states [compare Eq. (2.56)] as

$$\ln\tilde{M} \simeq \tfrac{1}{2}\int\frac{d^3p}{(2\pi)^3}\int d^3x \ln\frac{p^2+2\rho\beta\kappa^2}{p^2+2\rho\beta\kappa^2\cos(\varphi_c+\pi\theta_V)} . \tag{3.44}$$

The $\cos(\varphi_c+\pi\theta_V)$ can be approximated by 0 for $\vec{x}$ in the shell of thickness $(\pi/2)\lambda_D$ around ∂V, and by 1 if $\vec{x}$ is not in that region; thus

$$\ln\tilde{M} \simeq \tfrac{1}{2}S\frac{\pi}{2}\lambda_D\int\frac{d^3p}{(2\pi)^3}\ln\left(1+\frac{2\rho\beta\kappa^2}{p^2}\right). \tag{3.45}$$

This integral diverges linearly at large $|\vec{p}\,|$, so it is necessary to introduce an ultraviolet cutoff π/a, which would arise naturally in a lattice theory; the result is

$$\ln\tilde{M} = S\tfrac{1}{4}\rho\lambda_D\frac{\beta\kappa^2}{a} \tag{3.46}$$

apart from terms of higher order in a/λ_D, which are small if $\beta\kappa^2 \to 0$.

The origin of the ultraviolet divergence of the integral in Eq. (3.45) is the singularity of the Coulomb potential $1/|\vec{x}_\lambda - \vec{x}'_\lambda|$ at $\vec{x}_\lambda = \vec{x}'_\lambda$. If the potential were cut off at some small range r_0,

i.e., made finite for $|\vec{x}_\lambda - \vec{x}'_\lambda| < r_0$, then the high-momentum states with $|\vec{p}| \gtrsim \pi/r_0$ would be suppressed in the integral (3.45) and r_0 would replace a in Eq. (3.46). This singularity is associated with the *point* monopole field.

Comparison of Eqs. (3.43) and (3.46) shows that $\ln\tilde{M}$ is smaller than $S_V(\varphi_c)$ by a factor $\beta\kappa^2/a$, and this holds true for arbitrarily large V since both S_V and $\ln\tilde{M}$ are proportional to the surface area S. In the first approximation the factor $\tilde{M}$ can be ignored, so

$$M = e^{-S 2\pi\rho\lambda_D} . \tag{3.47}$$

As expected, the reduction of topological disorder associated with Debye screening leads to dependence of M on just the surface area S of ∂V, in contrast to the volume dependence (3.27) that occurs for a perfect gas.

The surface-area dependences $M \sim \exp(-S)$ of Eqs. (3.47) and (3.34) are similar in that both result from correlations in monopole positions. But they are different in that in the Coulomb gas (3.47) the correlation distance λ_D is dynamical while in the $m\bar{m}$ dipole gas (3.34) the correlation distance l is imposed by definition of the model.

Debye screening is a consequence of the long-range nature of the Coulomb potential. If the interaction of monopoles were a Yukawa potential, for instance, M would depend on the volume V as in Eq. (3.27) for V sufficiently large.

The expression (3.47) holds in the limit $R \gg \lambda_D$, where R is the linear size of V. More generally, it is expected that

$$M \sim \begin{cases} \exp(-R^2\rho\lambda_D) & \text{for } R \gg \lambda_D , \\ \exp(-R^3\rho) & \text{for } R \lesssim \lambda_D \end{cases}$$

apart from numerical factors. In the limit $\beta\kappa^2 \to 0$, which corresponds to the strong-coupling limit $g^2 \to \infty$ if the identification $\beta\kappa^2 = (2\pi/g)^2$ is made, the Debye length $\lambda_D \to \infty$ so the region over which the R^3 dependence holds becomes infinite. In this way the perfect-gas result is restored for $\beta\kappa^2 = 0$.

The loop integral W is expected to depend on the area $A(\Sigma)$ of Σ by the similarity of W and M. The quantities W and M measure the magnetic flux through the surfaces Σ and ∂V. In the limit in which Σ and ∂V become infinite, the distinction that Σ is an open set while ∂V is closed should become irrelevant. Then W should depend on the area $A(\Sigma)$, in accordance with the dependence (3.47) of M on the surface area of ∂V, aside from effects of the boundary $\partial\Sigma$ of Σ which should be proportional to $L(\partial\Sigma)$.

Polyakov described the dependence of W on Σ in Ref. 10. In analogy with Eq. (3.41) W may be estimated as

$$W = e^{-S_\Sigma(\varphi_c)}, \tag{3.48}$$

where φ_c is now the field that minimizes $S_\Sigma(\varphi)$; the factor $\tilde{W}$ due to small fluctuations of φ away from φ_c is neglected. The field φ_c is not known to me but its qualitative behavior can be guessed from the following arguments. The second term in Eq. (3.39b) for S_Σ is minimized by setting $\varphi = -\Phi_\Sigma$. Now $\Phi_\Sigma(\vec{x})$ is discontinuous on Σ: the discontinuity of Φ_Σ across the $x_3 = 0$ plane is [see Eqs. (3.9) and (3.10)]

$$\lim_{x_3 \to 0}[\Phi_\Sigma(x_\perp, x_3) - \Phi_\Sigma(x_\perp, -x_3)] = 2\pi\theta_\Sigma(x_\perp),$$

where $\theta_\Sigma(x_\perp)$ is equal to 1 if $x_\perp \in \Sigma$, and to 0 if $x_\perp \notin \Sigma$. Thus, for the field $\varphi = -\Phi_\Sigma$ the first term in S_Σ is infinite because of the singularity of $\nabla\Phi_\Sigma$ at Σ; the action is proportional to $A(\Sigma)\Lambda/\beta\kappa^2$, where Λ is a momentum-space ultraviolet cutoff. On the other hand, the first term in S_Σ is minimized by setting $\varphi = 0$. For this field the action is proportional to ρR^3 where R is the radius of Σ. The full action S_Σ is minimized by a field φ_c that balances these two terms. φ_c is roughly equal to $-\Phi_\Sigma$ but with the discontinuity across Σ smoothed out over a distance comparable to the Debye length λ_D. For such a field the ultraviolet cutoff Λ is not needed and is effectively replaced by λ_D^{-1}. These remarks, and the more explicit treatment due to Polyakov,[10] lead to the result

$$S_\Sigma(\varphi_c) = C\rho\lambda_D A(\Sigma), \tag{3.49}$$

where C is a numerical constant and terms proportional to the perimeter $L(\partial\Sigma)$ are ignored. Thus W obeys an area law in this system.

The result (3.49) can be illustrated qualitatively by considering instead of S_Σ the approximation $S_\Sigma^{(2)}$ obtained by expanding the cosine in Eq. (3.39b):

$$S_\Sigma^{(2)} = \int d^3x \left[\frac{1}{2\beta\kappa^2}(\nabla\varphi)^2 + \rho(\varphi + \Phi_\Sigma)^2\right]. \tag{3.50}$$

$S_\Sigma^{(2)}$ is minimized by the field

$$\varphi_c^{(2)}(\vec{x}) = -\int \frac{d^3p}{(2\pi)^3} e^{i\vec{p}\cdot\vec{x}} \hat{\Phi}_\Sigma(\vec{p}) \frac{2\lambda_D^{-2}}{\vec{p}^2 + 2\lambda_D^{-2}} , \tag{3.51}$$

where $\hat{\Phi}_\Sigma(\vec{p})$ is the Fourier transform of $\Phi_\Sigma(\vec{x})$. Thus $\varphi_c^{(2)}$ is roughly equal to $-\Phi_\Sigma$ but with the high-momentum components, those with $\vec{p}^2 \gtrsim \lambda_D^{-2}$, suppressed; in particular, the discontinuity on Σ is smoothed out. The minimum of $S_\Sigma^{(2)}$ is

$$S_\Sigma^{(2)}(\varphi_c^{(2)}) = \rho \int \frac{d^3p}{(2\pi)^3} |\hat{\Phi}_\Sigma(\vec{p})|^2 \frac{\vec{p}^2}{\vec{p}^2 + 2\lambda_D^{-2}} . \tag{3.52}$$

It can be shown that in the limit $\Sigma \to \infty$, or equivalently $\lambda_D \to 0$, $S_\Sigma^{(2)}(\varphi_c^{(2)})$ is asymptotically

$$S_\Sigma^{(2)}(\varphi_c^{(2)}) \sim \sqrt{2}\pi^2\rho\lambda_D A(\Sigma), \tag{3.53}$$

which illustrates the dependence on $A(\Sigma)$ in Eq.

(3.49).

The area law (3.49) shows that the correlations measured by W are intermediate between the two earlier cases, (3.25) and (3.32).

The result (3.49) holds for Σ large compared to λ_D; more generally,

$$W \sim \begin{cases} e^{-\rho\lambda_D R^2} & \text{for } R \gg \lambda_D , \\ e^{-\rho R^3} & \text{for } R \leq \lambda_D , \end{cases}$$

where R is the radius of Σ. Thus Debye screening produces the area-law behavior associated with linear confinement. In the limit $\beta \to 0$ the Debye length $\lambda_D \to \infty$ and the perfect-gas behavior (3.25) is recovered.

The loop integral W does not depend directly on 3D monopoles, but rather on the vorticity in the 2D subspace $x_3 = 0$. The Debye length λ_D is the correlation length of fluctuations of magnetic charge. If the region Σ is large compared to λ_D then W sees effectively a gas of uncorrelated point vortices with 2D density $\xi = \rho\lambda_D$; these are the vortices created by monopoles within a distance λ_D in the x_3 direction from Σ. This leads to the area law $W \sim \exp(-\xi R^2) = \exp(-\rho\lambda_D R^2)$.

The final topic to be discussed in this section is the definition of dual parameters for the quantities W and M. For simplicity the dual parameters will be described for the Abelian gauge theory. The discussion is formulated in the operator-field theory with the gauge choice $A_0 = 0$, the temporal gauge.

The operator dual to the loop integral $\exp(ig \oint_{\partial\Sigma} \times d\vec{s} \cdot \vec{A})$ is constructed in the following way.[21,34] Let $|\vec{A}\rangle$ be the eigenstate of the gauge-field operator with eigenvalue $\vec{A}(\vec{x})$. The dual operator U, which depends on a simple closed curve C, is defined by its action on $|\vec{A}\rangle$:

$$U|\vec{A}\rangle = |\vec{A} + \vec{A}_c\rangle , \tag{3.54}$$

where $\vec{A}_c(\vec{x})$ is the vector potential of a magnetic flux loop around C with flux π/g. That is, $\vec{A}_c(\vec{x})$ has the property that for any closed curve C',

$$\oint_{C'} d\vec{l} \cdot \vec{A}_c = \frac{\pi}{g} n(C', C) , \tag{3.55}$$

where $n(C', C)$ is the number of times C' winds around C; in particular, if C' and C are planar then

$$n(C', C) = \begin{cases} 1 & \text{if } C' \text{ and } C \text{ are linked,} \\ 0 & \text{if not.} \end{cases} \tag{3.56}$$

An explicit formula for $\vec{A}_c(\vec{x})$ is

$$\vec{A}_c(\vec{x}) = \frac{\pi}{g} \nabla \times \oint_C d\vec{l} \, \frac{1}{4\pi(\vec{x} - \vec{l})} ; \tag{3.57}$$

the corresponding magnetic field is

$$\nabla \times \vec{A}_c(\vec{x}) = \frac{\pi}{g} \oint_C d\vec{l} \, \delta^3(\vec{x} - \vec{l}) . \tag{3.58}$$

Since the magnetic field $\nabla \times \vec{A}_c$ vanishes except on the curve C, the vector potential $\vec{A}_c$ is a pure gauge field singular on C. That is, $\vec{A}_c = \nabla\Lambda_c$ where the gauge function $\Lambda_c(\vec{x})$ is multivalued and ill-defined on C. As an explicit example, suppose C is the x_3 axis; then the gauge function is $\Lambda_c(\vec{x}) = \pi\varphi/g$, where the polar coordinates of $\vec{x}$ are (r, θ, φ).

It follows from Eq. (3.54) that the operator U transforms $\vec{A}$ by the singular gauge transformation Λ_c:

$$U^\dagger \vec{A} U = \vec{A} + \nabla\Lambda_c . \tag{3.59}$$

Since U is a shift operator of the gauge field $\vec{A}$, it can be written as

$$U = \exp\left[i \int d^3x \, \vec{E}(\vec{x}) \cdot \vec{A}_c(\vec{x}) \right], \tag{3.60}$$

where $\vec{E}$ is the canonical momentum of $\vec{A}$. Thus U is the creation operator of a loop of magnetic flux around C.

The interesting point is the commutation relation of U and the loop integral:

$$U^\dagger W U = e^{i\pi n(\partial\Sigma, C)} W , \tag{3.61a}$$

where

$$W = \exp\left(ig \oint_{\partial\Sigma} d\vec{s} \cdot \vec{A} \right) . \tag{3.61b}$$

If the curves C and $\partial\Sigma$ are not linked then U commutes with the loop integral; if C and $\partial\Sigma$ are linked then these operators anticommute. The commutation relation (3.61) justifies calling U the dual operator of the loop integral W. This duality involves also intercharge of electric and magnetic fields because W creates a loop $\partial\Sigma$ of electric flux while U creates a loop C of magnetic flux.

't Hooft has shown how to generalize the operator U to a non-Abelian gauge theory.[21,34] Again U is defined as the operator that transforms $\vec{A}_a$ by a gauge transformation singular along a closed curve C; and it transforms the loop integral $P \exp(i \oint_{\partial\Sigma} ds \cdot A)$ according to a relation analogous to Eq. (3.61).

The implications of the relation (3.61) have been discussed by 't Hooft. It is argued that this relation implies the lack of declustering $\langle 0|WU|0\rangle \neq \langle 0|W|0\rangle\langle 0|U|0\rangle$ even in the limit of infinite separation of the curves $\partial\Sigma$ and C. The failure of declustering can result from the existence of massless excitations, photons or gluons; but if there are no massless excitations then one or both of the expectation values $\langle 0|W|0\rangle$ and $\langle 0|U|0\rangle$ should decrease according to an area law $\exp(-A)$ where A = area enclosed by $\partial\Sigma$ or C.

The dual operator U can thus also be used to distinguish between topological disorder and order: if $\langle 0|U|0\rangle$ does not decrease as $\exp[-A(C)]$ then the loop integral W must obey an area law, indicating topological disorder. The connection is that U creates a loop of magnetic flux when it acts on the vacuum state $|0\rangle$. If the overlap of $U|0\rangle$ with $|0\rangle$, that is $\langle 0|U|0\rangle$, is not small, then evidently the vacuum is not changed much by the presence of the additional flux loop C. Then there must be topological disorder, so the loop integral W is small. To make this connection more explicit, consider the path-integral formula for $\langle 0|U|0\rangle$:

$$\langle 0|U|0\rangle = \int dA^i(\vec{x}) \int_{P(A,A+A_c)} dA'^{\mu}(x) e^{-S(A')},$$

where $\int_{P(A,A+A_c)}$ indicates integration over the set of paths $A'_\mu(x)$ that begin at $A^i(\vec{x})$ at Euclidean time $t=-\infty$ and end at $A^i(\vec{x})+A^i_c(\vec{x})$ at time $t=+\infty$. If $\langle 0|U|0\rangle$ is not small, then there is a contribution to the path integral from paths in which a magnetic flux loop is created. Those paths produce vortices in the effective 2D theory $P_2(\varphi^\alpha)$ [Eq. (2.1)] because a magnetic flux loop sliced by a 2D plane creates on the plane a point vortex and antivortex separated by a distance of the order of the size of C. The flux π/g implies that the vortices satisfy the quantization condition (2.5). Thus a large value of $\langle 0|U|0\rangle$ implies topological disorder in the effective 2D theory and area-law behavior of W.

Finally, an operator $U(\vec{c})$ dual to the monopole operator M is the operator that creates a monopole field centered at $\vec{c}$:

$$U(\vec{c}) = \exp\left(i \int d^3x \vec{E}(\vec{x}) \cdot \vec{A}_m(\vec{x}-\vec{c})\right). \tag{3.62}$$

The commutation relation analogous to Eq. (3.61) is

$$U^\dagger(\vec{c}) M U(\vec{c}) = e^{i\vartheta v(\vec{c})} M, \tag{3.63a}$$

where

$$M = \exp\left(\frac{i}{2} g \int'_{\partial V} d\vec{\sigma} \cdot \nabla \times \vec{A}\right) \tag{3.63b}$$

and $\int'_{\partial V}$ is the surface integral over ∂V *minus* infinitesimal surfaces around every Dirac string. The operator $U(\vec{c})$ can also distinguish between states of topological order and disorder. For instance, if the vacuum contains uncorrelated monopoles then the state obtained by creating an additional m and $\bar{m}$ at $\vec{c}$ and $\vec{c}'$, namely $U(\vec{c})U^\dagger(\vec{c}')|0\rangle$, has a nonzero overlap with $|0\rangle$ that does not decrease rapidly as the distance $|\vec{c}-\vec{c}'|\to\infty$; but if there is no topological disorder then $\langle 0|U(\vec{c}) \times U^\dagger(\vec{c}')|0\rangle$ decreases rapidly as $|\vec{c}-\vec{c}'|\to\infty$.

Although $U(\vec{c})$ is defined as a dual operator for M, its commutation relation with W is also relevant:

$$U^\dagger(\vec{c}) W U(\vec{c}) = e^{i\Phi_\Sigma(\vec{c})} W. \tag{3.64}$$

Both of the dual operators U and $U(\vec{c})$ *create* magnetic flux and thus have nontrivial commutation relations with the parameter that *measures* magnetic flux W.

IV. MONOPOLE IONIZATION

In Sec. III the influence of monopole vacuum fields is discussed in terms of some simple models of a gas of monopoles. In these models the Wilson loop integral W indicates the topological disorder created by the monopoles. The purpose of the present section is to review some ideas about the origin of monopole field configurations in Abelian and non-Abelian gauge theories.[16,13,14]

The path-integral formula Eq. (3.1) for the vacuum functional shows how monopole vacuum fields are created by paths $A'^i_a(\vec{x}, x_4)$ in the space of field configurations that are monopole fields at time $x_4=0$. Two examples of such paths will be considered here: monopole loops[16] and, in a non-Abelian gauge theory, merons.[13] The contribution of these paths will be discussed in terms of models obtained by analogy with instanton calculations.[10] These models imply a transition from bound $m\bar{m}$ pairs to free monopoles in the effective 3D theory described by the vacuum functional, the monopole ionization referred to in the title. The interesting question is how these two mechanisms of production of monopoles differ.

The discussion of the monopole loop applies to an Abelian gauge theory. The theory must be constructed as a lattice gauge theory to avoid ultraviolet divergences. However, the topological field configurations being considered can sensibly be treated in a continuum approximation because they spread over many lattice sites and vary slowly compared to the lattice spacing a. For my purposes the details of the construction of a lattice theory can be ignored; the finite lattice spacing is mainly needed just to provide an ultraviolet cutoff for otherwise divergent integrals.

The monopole loop is a path in the space of field configurations for which a monopole and antimonopole are created at some time τ_i, separate along trajectories $\vec{c}(x_4)$ and $\vec{c}'(x_4)$, and annihilate at a later time τ_f. The creation and annihilation occur at zero separation so $\vec{c}(x_4)=\vec{c}'(x_4)$ at $x_4=\tau_i$ and τ_f. The field corresponding to this process is the monopole loop

$$A^i(\vec{x}, x_4) = A^i_m(\vec{x}-\vec{c}(x_4)) - A^i_m(\vec{x}-\vec{c}'(x_4)), \tag{4.1}$$

where A^i_m is the Dirac monopole field (3.5). This field creates an $m\bar{m}$ pair in the 3D subspace at

$x_4=0$, with m and $\bar{m}$ positions at $\vec{c}(0)$ and $\vec{c}'(0)$. If the contribution of monopole-loop paths to the path integral is limited to loops of characteristic size l, then the resulting $m\bar{m}$ pairs will have separations $\lesssim l$. Free monopoles will exist if arbitrarily large monopole loops contribute to the path integral.

The process described in the previous paragraph can also be viewed as a magnetic current loop C in 4D space. Let C be a loop defined by the equation $x_\mu = z_\mu(\theta)$ where θ is an angular parameter, $0 \leq \theta \leq 2\pi$ and $z_\mu(0) = z_\mu(2\pi)$. The curve $z^\mu(\theta)$ is related to the m and $\bar{m}$ trajectories $\vec{c}(x_4)$ and $\vec{c}'(x_4)$ in the following way: Let $\theta(x_4)$ and $\bar{\theta}(x_4)$ be the two solutions of the equation $x_4 = z_4(\theta)$; then the trajectories are $\vec{c}(x_4) = \vec{z}[\theta(x_4)]$ and $\vec{c}'(x_4) = \vec{z}[\bar{\theta}(x_4)]$.

The curve C is a loop of *magnetic* current. The current density is

$$k_\mu(x) = \kappa \oint_C dz_\mu \delta^4(z-x), \tag{4.2}$$

where $\kappa = 2\pi/g$ is the monopole charge. The charge density at time x_4 is

$$k_0(\vec{x}, x_4) = \kappa\theta(\tau_f - x_4)\theta(x_4 - \tau_i) \times[\delta^3(\vec{x} - \vec{c}(x_4)) - \delta^3(\vec{x} - \vec{c}'(x_4))]. \tag{4.3}$$

The fields associated with $k_\mu(x)$ are defined by

$$\begin{aligned} &\partial_\nu \tilde{F}_{\mu\nu} = k_\mu, \\ &\tilde{F}_{\mu\nu} = \tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}F_{\alpha\beta}. \end{aligned} \tag{4.4}$$

$\tilde{F}_{\mu\nu}$ is the dual of the electromagnetic field tensor $F_{\mu\nu}$ since k_μ is a magnetic current density.

The obvious way to solve Eq. (4.4) is to introduce a dual field $\bar{A}_\mu$ defined by

$$\bar{A}_\mu(x) = \int d^4y D(x-y) k_\mu(y), \tag{4.5}$$

where $D(x)$ is the inverse of the 4D Laplacian

$$-\partial_\mu\partial_\mu D(x) = \delta^4(x), \tag{4.6}$$

and write $\tilde{F}_{\mu\nu}$ as

$$\tilde{F}_{\mu\nu} = \partial_\mu \bar{A}_\nu - \partial_\nu \bar{A}_\mu. \tag{4.7}$$

However, the path integral (3.1) is an integral over the gauge field $A_\mu(x)$ not the dual field $\bar{A}_\mu(x)$. Thus it is necessary to construct a gauge field of the magnetic charge following Dirac.[35] The construction involves introduction of a singular surface Δ, the surface swept out by the Dirac string as the monopole moves along C; the boundary of Δ is C. Let Δ be defined by the equation $x_\mu = y_\mu(\theta, \xi)$ where $0 \leq \theta \leq 2\pi$ and $0 \leq \xi < \infty$; the boundary of Δ is at $\xi = 0$ so $y^\mu(\theta, 0) = z^\mu(\theta)$. The Dirac monopole field is

$$A_\mu(x) = \kappa \int_\Delta d\sigma_{\alpha\beta}\epsilon_{\mu\nu\alpha\beta}\partial_\nu D[x - y(\sigma)]. \tag{4.8}$$

The corresponding field tensor is

$$\begin{aligned} F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu \\ &= \tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}\tilde{F}_{\alpha\beta} + G_{\mu\nu}, \end{aligned} \tag{4.9}$$

where $\tilde{F}_{\alpha\beta}$ obeys Eq. (4.4) and $G_{\mu\nu}$ is a singular field that is nonzero only on Δ:

$$G_{\mu\nu}(x) = \kappa \int_\Delta d\sigma_{\mu\nu}\delta^4[x - y(\sigma)]. \tag{4.10}$$

To see the connection between the monopole loop field $A_\mu(x)$ and the 3D monopole field $A^i_m(\vec{x})$ in Eq. (3.5), consider an infinite loop C defined by the trajectories of two static monopoles located at $\vec{c} = 0$ and $\vec{c}' = (0, 0, l)$. The surface Δ is an infinite strip in the (x_3, x_4) plane defined by $-\infty \leq x_4 \leq +\infty$ and $0 \leq x_3 \leq l$. The corresponding Dirac field $A_\mu(x)$ in Eq. (4.8) is

$$\begin{aligned} &A_0(x) = 0, \\ &A_i(x) = A^i_m(\vec{x}) - A^i_m(\vec{x} - \vec{c}'). \end{aligned} \tag{4.11}$$

Thus the 4D monopole loop creates a 3D $m\bar{m}$ pair with separation comparable to the loop size.

The position of the singular sheet Δ can be changed by a singular gauge transformation of $A_\mu(x)$. Since Δ is not gauge invariant it should be unobservable. The magnetic charge obeys the Dirac quantization condition $\kappa = 2\pi/g$. This implies that the singular field $G_{\mu\nu}(x)$ is unobservable in a lattice gauge theory in which $A_\mu(x)$ appears only as $\exp(ig\int ds \cdot A)$ because then $G_{\mu\nu}$ occurs only as $\exp(igG) = \exp(i2\pi)$.

The contribution of the monopole-loop field (4.8) to the path integral depends on the action of this configuration. If the action is defined as in Eq. (1.3) as the integral of F^2, then it is infinite because of the singular sheet Δ. However, it can be argued that the unobservability of Δ implies that the action appropriate for monopole fields is the Dirac action[35]

$$\begin{aligned} S &= \tfrac{1}{4}\int d^4x (F_{\mu\nu} - G_{\mu\nu})(F_{\mu\nu} - G_{\mu\nu}) \\ &= \tfrac{1}{4}\int d^4x \tilde{F}_{\mu\nu}\tilde{F}_{\mu\nu} \end{aligned} \tag{4.12}$$

for which the singular field $G_{\mu\nu}$ is removed. Peskin has shown quite explicitly how the Dirac action arises for monopole loops in a lattice gauge theory.[16]

The action (4.12) is

$$S = \tfrac{1}{2}\int d^4x\, d^4y\, k_\mu(x) k_\mu(y) D(x-y), \tag{4.13}$$

or, by Eq. (4.2) for $k_\mu(x)$ the action is the double

line integral

$$S=\tfrac{1}{2}\kappa^2\oint_c dz^\mu\oint_c dz'^\mu D(z-z'). \tag{4.14}$$

The Green's function $D(x)$ of the 4D Laplacian is

$$D(x)=\frac{1}{4\pi^2x^2}. \tag{4.15}$$

The integral (4.14) is divergent because of the singularity of $D(z-z')$ at $z=z'$. It must be rendered finite by excluding the small segment of C' of points z' such that $|z-z'|\leq\epsilon$. Then the action S is the sum of a divergent term S_0 and a term S_c that is finite as $\epsilon\to 0$,

$$S=S_0+S_c. \tag{4.16a}$$

The divergent term is

$$S_0=\frac{\kappa^2}{4\pi^2}\frac{L}{\epsilon}, \tag{4.16b}$$

where L is the length of the curve C. The finite term is, in the limit $\epsilon\to 0$,

$$S_c=\frac{\kappa^2}{4\pi^2}\oint_c dz_i\oint_c dz'_j\frac{\hat{n}(\vec{z})\cdot(\vec{z}-\vec{z}')}{|\vec{z}-\vec{z}'|^4}[\delta_{ij}\hat{n}(\vec{z})\cdot(\vec{z}'-\vec{z})-n_j(\vec{z})(z'-z)_i], \tag{4.16c}$$

where $\hat{n}(\vec{z})$ is the unit outward normal to the curve C at $\vec{z}$. S_c is dimensionless and finite; if C is a circle then $S_c=-\kappa^2/4$. In general, S_c depends on the shape of the curve C but not its overall scale L.

The L-dependent term S_0 depends on the cutoff ϵ. This shows the necessity of an ultraviolet cutoff. In a lattice theory a cutoff ϵ would occur naturally and would be of the order of the lattice spacing a; a rigorous treatment of the lattice theory is needed to identify the correct value of ϵ, but eventually I shall simply set ϵ equal to a. In the continuum treatment, the divergence of S as $\epsilon\to 0$ derives from the fact that the monopole field has a *point* singularity at the position of the monopole center. The divergence could be eliminated by smearing out the source of the monopole field $k_\mu(x)$ into a charge of finite extent.

The contribution of monopole loops to the partition function Z will be discussed under the assumption that the path integral (1.4) is dominated by field configurations $A_\mu(x)$ of the form

$$A_\mu(x)=\bar{A}_\mu(x;\lambda_1,\ldots,\lambda_n)+\varphi_\mu(x), \tag{4.17}$$

where $\bar{A}_\mu(x;\lambda_1,\ldots,\lambda_n)$ is a multiloop field with λ_i the collective coordinates needed to specify the ith loop, for which the fluctuation φ_μ of A_μ away from $\bar{A}_\mu$ is small. Then the path integral $\int dA_\mu(x)$ can be replaced by integrals over the collective coordinates $\int\pi_i\,d\lambda_i$ and small fluctuations $\int d\varphi_\mu(x)$. The action can be written

$$S(A_\mu)=S(\lambda_1,\ldots,\lambda_n)+S(\varphi_\mu), \tag{4.18}$$

where $S(\lambda_1,\ldots,\lambda_n)$ is the action of the multiloop field $\bar{A}_\mu$; the term linear in φ_μ vanishes because $\bar{A}_\mu$ obeys the Euclidean field equations $\partial_\nu F_{\mu\nu}=0$, and $S(\varphi_\mu)$ is the quadratic action (1.3) since φ_μ is small. The small fluctuations φ_μ do not interact with the background field $\bar{A}_\mu(x)$ because of the quadratic form of the action; physically this corresponds to the fact that photons, which are the quanta of the small fluctuations $\varphi_\mu(x)$, are not charged. This separation of A_μ into a monopole loop plus small fluctuation leads to factorization of the partition function: $Z=Z_0Z_m$ where Z_0 depends only on the fluctuations φ_μ and Z_m is the partition function of a system of monopole loops

$$Z_m=\sum_n\frac{1}{n!}\int d\lambda_1\cdots d\lambda_n e^{-S(\lambda_1,\ldots,\lambda_n)}f(\lambda_1,\ldots,\lambda_n), \tag{4.19}$$

where $f(\lambda_1,\ldots,\lambda_n)$ is the volume in the space of fields $\varphi_\mu(x)$ of fluctuations that are identical to fluctuations of the collective coordinates λ_i.

To proceed it is necessary to assume that the monopole loops are independent. That is, interactions between elements of different loops are ignored and the action $S(\lambda_1,\ldots,\lambda_n)$ is taken to be the sum of the actions of individual loops

$$S(\lambda_1,\ldots,\lambda_n)=\sum_i S(\lambda_i). \tag{4.20}$$

Further, the factor $f(\lambda_1,\ldots,\lambda_n)$ is taken to be

$$f(\lambda_1,\ldots,\lambda_n)=\prod_i f(\lambda_i). \tag{4.21}$$

These approximations are valid if the partition sum is dominated by configurations with a small density of monopole loops. Then Z_m becomes

$$Z_m=\exp\left(\int d\lambda\, e^{-S(\lambda)}f(\lambda)\right). \tag{4.22}$$

The action $S(\lambda)$ is the sum of a term $S_0[l(\lambda)]$ that depends only on the length $l(\lambda)$ of the loop λ and a convergent term $S_c(\lambda)$. Thus Z_m can be rewritten

$$Z_m=\exp\left(\frac{1}{a}\int_a^\infty dl\, e^{-S_0(l)}P(l)\right), \tag{4.23}$$

where $P(l)$ is the distribution in length

$$P(l)=\int d\lambda\, e^{-S_c(\lambda)}f(\lambda)a\delta[l-l(\lambda)]. \tag{4.24}$$

The lower limit of integration of the length l is identified in Eq. (4.23) as just the lattice spacing a; this is somewhat ambiguous, but since it is the large l limit that is of most interest, the ambiguity is unimportant.

The distribution $P(l)$ is essentially just the total number of loops of length l; this is modified slightly by the quantity $e^{-S_0 l}$ in Eq. (4.24). The loops are confined to a lattice. The number of links in a loop of length l is l/a. Since the loop can be located at any point in the 4D space, $P(l)$ can be written

$$P(l)=\frac{\Omega}{a^4}p(l), \tag{4.25}$$

where Ω is the volume of four-space and $p(l)$ is the number of loops of length l constrained to pass through the origin. The number of curves that begin at the origin and have l/a links is $(2d-1)^{l/a}$ where $d=4$ is the number of dimensions. Not all of these curves are loops, but the dominant factor in $p(l)$ should be $(7)^{l/a}$. Thus it may be estimated that $p(l)$ is of the form[36]

$$p(l)=(7)^{l/a}h(l), \tag{4.26}$$

where the factor $h(l)$, which derives from the requirement that the curve return to the origin, is less important than the first factor. The importance of the lattice spacing a in the treatment of monopole loops is evident in Eq. (4.26).

The partition function becomes

$$Z_m=\exp\left\{a^{-5}\Omega\int_a^\infty dl\,h(l)\exp\left[-\left(\frac{\beta\kappa^2}{4\pi^2}-\ln 7\right)\frac{l}{a}\right]\right\}. \tag{4.27}$$

Here the ultraviolet cutoff ϵ in S_0 is identified with the lattice spacing a. This choice is uncertain *and affects the estimate of the phase-transition point given below.* More generally, it should only be said that ϵ is *of order* a.

The partition function Z_m of the monopole sector obviously predicts a phase transition. At large $\beta\kappa^2$ the exponential in the integrand of (4.27) cuts off loops at a size comparable to $a(\beta\kappa^2/4\pi^2-\ln 7)^{-1}$. At sufficiently small $\beta\kappa^2$ the integral over l is unbounded and the approximations leading to Eq. (4.27) break down; presumably then loops of arbitrarily large length l contribute significantly to Z_m. The expression (4.27) for Z_m implies that the critical value of $\beta\kappa^2$ where the phase transition occurs is

$$\frac{\beta\kappa^2}{4\pi^2}=\ln 7 \tag{4.28}$$

The magnetic charge is $\kappa=2\pi/g$ and β can be absorbed into the definition of g, so the identification is made $g^{-2}=\beta\kappa^2/4\pi^2$; the critical value of g^2 is then g_c^2:

$$\frac{1}{g_c^2}=\ln 7. \tag{4.29}$$

This estimate is subject to several ambiguities that could change the value of g_c^2, though not the fact that there is a phase transition. The obvious ambiguity is the treatment of the cutoff ϵ in S_0; Eq. (4.29) was obtained with the simple choice $\epsilon=a$. Other possible sources of error in (4.29) are the assumption that $h(l)$ is dominated by $\exp(l)$ in Eq. (4.27), which might be invalid because of the contribution of rapidly fluctuating loops for which (4.16) is a poor approximation of the action; and the assumption that interactions between elements of different loops can be neglected.

If the unknown factor $h(l)$ in Eq. (4.27) is ignored, then Z_m can be evaluated explicitly. An interesting derived quantity is the specific heat C. For g^2 near g_c^2 and $g^2<g_c^2$, and for the infinite-volume limit of the system, the specific heat is

$$C=\frac{2}{a^4}g^2\left(1-\frac{g^2}{g_c^2}\right)^{-3}\exp\left(-\frac{1}{g^2}+\frac{1}{g_c^2}\right). \tag{4.30}$$

Here again the identification $g^{-2}=\beta\kappa^2/4\pi^2$ has been made. Thus the specific heat of this system diverges at the transition point $g^2=g_c^2$. In contrast, Eq. (2.35) shows that the specific heat is not singular at the vortex ionization point $\beta_c=8\pi/\kappa^2$ for the 2D Coulomb gas of vortices. The reason for this difference is that in Eq. (4.27) for Z_m the integrand tends to infinity *exponentially* as $l\to\infty$ if $g^2>g_c^2$; thus all thermodynamic quantities diverge at $g^2=g_c^2$. In contrast, in Eq. (2.33) for the partition function of the 2D Coulomb gas, the integrand diverges only as a *power* as $|x-y|\to\infty$ since the potential $G(x-y)$ is logarithmic; thus some quantities, e.g., the specific heat, remain finite at the transition point β_c.

The monopole-loop phase transition described above implies a monopole-ionization transition in the effective 3D theory defined by the vacuum functional $P_3(A_a^i)$. For $g^2<g_c^2$ only monopole loops with size of order a contribute to the path integral; these create in the 3D subspace $m\bar{m}$ dipoles with separation of order a. The short-range dipole fields do not disorder the system, so the loop integral W has the perimeter-law behavior of a nonconfining theory. For $g^2>g_c^2$ arbitrarily large monopole loops contribute, and these create free monopoles in the 3D theory. In this phase W may decrease more rapidly as $\Sigma\to\infty$ than a perimeter law; e.g., W may have the area-law behavior of a confining theory.

The monopole ionization can also be discussed

in terms of the dependence of the quantity M on the region V in 3D subspace. M is the expectation value of $\exp(i\pi n_V)$ where n_V is the number of monopoles in V [see Eq. (3.20)]. A monopole loop affects this quantity if of the two points of intersection of the loop with the 3D subspace $x_4=0$ one is inside V. In the phase with $g^2<g_c^2$, only loops near the surface ∂V of V affect M; then $M\sim\exp(-S)$ where S is the area of ∂V. In the phase with $g^2>g_c^2$, there are arbitrarily large loops and these need not lie near ∂V in order to affect V; then M may have a different dependence on V than in the ordered phase.

The results of Sec. III show that the dependence of W on Σ and of M on V in the large-g^2 phase depends on the density of monopoles and, more importantly, on their effective interactions. It appears to me that in order to determine these quantities, it is necessary to go beyond the simplest estimates, Eqs. (4.16) and (4.26), of the action $S(\lambda)$ and distribution of loops $P(\lambda)$; the nondominant terms like $S_c(\lambda)$ and $h(l)$ are needed.

For a lattice gauge theory, the relevance of the monopole sector may be rather limited in the disordered phase $g^2>g_c^2$. For g^2 sufficiently large, the strong-coupling expansion is presumably valid. The lowest-order approximation to the strong-coupling expansion $g^2=\infty$ corresponds to complete disorder: for $\beta=0$ *all* field configurations in the partition sum are equally probable. Then the partition sum is no longer adequately described as a sum over just monopole loops and small Gaussian fluctuations. The monopole sector is applicable for only a limited range of g^2 when $g^2>g_c^2$.

This picture of monopole ionization is not inconsistent with the results of the Monte Carlo study of the Abelian lattice gauge theory of Creutz.[18] Specifically, he finds a phase transition at $g_c^2\simeq 1$ such that (1) the loop integral W changes from perimeter law at $g^2<g_c^2$ to area law at $g^2>g_c^2$, and (2) the specific heat C is apparently singular at g_c^2 since the average energy E does not converge to a single value at that point.

This naive description of the contribution of monopole loops to the path integral is meant to be heuristic. The basic idea is the separation of field configurations into monopole loops and small fluctuations. To make an exact separation of this kind requires a precise treatment of the lattice, to account correctly for large-momentum fluctuations. To be specific, a monopole loop for which the curve C is rapidly varying might be described alternatively as a loop field for which C is smooth plus large-momentum fluctuations. However, the precise treatment of these ultraviolet effects probably does not affect the nature of the phase transition, which involves the large loops.

The monopole-loop field is the path in field-configuration space that creates a 3D monopole pair appropriate to an Abelian gauge theory. This field can obviously be embedded in a non-Abelian gauge theory. However, in a non-Abelian theory there is another kind of path that creates a 3D monopole, namely, the meron configuration. Meron configurations are solutions to the Euclidean gauge-field equations, except at singular points.[37] Their role in the path integral has been discussed in Refs. 13 and 14.

The meron configuration in an SU(2) gauge theory can be written in the temporal gauge, i.e., the gauge choice $A_a^4=0$, as

$$A_a^i(\vec{x},x_4)=\frac{1}{g}\epsilon_{aij}\frac{x_j}{\vec{x}^2}\left[1+\frac{x_4}{(\vec{x}^2+x_4^2)^{1/2}}\right]. \tag{4.31}$$

This field tends to zero as $x_4\to-\infty$ and to a pure gauge field as $x_4\to\infty$. At $x_4=0$ it is precisely the Wu-Yang monopole field $A_{ma}^i(\vec{x})$ in Eq. (3.16).

The meron in Eq. (4.31) is centered at $\vec{x}=0$ and $x_4=0$. More generally, a meron centered at $\vec{x}=\vec{c}$ and $x_4=c_4$ creates the following monopolelike field in the 3D subspace $x_4=0$:

$$A_a^i(\vec{x}-\vec{c},-c_4)=\frac{1}{g}\epsilon_{aij}\frac{(x-c)_j}{(\vec{x}-\vec{c})^2}\left[1-\frac{c_4}{[(\vec{x}-\vec{c})^2+c_4^2]^{1/2}}\right]. \tag{4.32}$$

This field is similar to the Wu-Yang field $A_{ma}^i(\vec{x}-\vec{c})$ centered at $\vec{x}=\vec{c}$. However, it depends on the additional parameter c_4 which has the character of a *scale* parameter: A_a^i approaches the monopole field in the limit $|\vec{x}-\vec{c}|\gg c_4$.

The action of the meron configuration is logarithmically divergent because of both the singularity at $x=0$ and the slow decrease at large distances. The point singularity at $x=0$ is irrelevant; it could be smoothed out in a continuum formulation of the theory, and it is eliminated by the finite lattice spacing. But the slow decrease of $A_a^i(\vec{x},x_4)$, i.e., as $|\vec{x}|^{-1}$ as $|\vec{x}|\to\infty$, is crucial if the meron is to produce a long-range 3D monopole field. The divergence of the action implies that the single meron configuration does not contribute to the path integral.

The meron-pair field configuration,[37] which creates a pair of monopoles in the 3D subspace, has finite action. If the meron centers are both located on the x_4 axis, say at $c=(\vec{0},T)$ and $c'=(\vec{0},T')$, then the meron-pair configuration can be written in the temporal gauge ($A_a^4=0$) as

$$A_a^i(\vec{x},x_4)=\frac{1}{g}\epsilon_{aij}\frac{x_j}{\vec{x}^2}\left\{1-\frac{\vec{x}^2+(x_4-T)(x_4-T')}{[\vec{x}^2+(x_4-T)^2]^{1/2}[\vec{x}^2+(x_4-T')^2]^{1/2}}\right\}. \tag{4.33}$$

If the separation $|T-T'|$ of the merons is large then for x_4 near T or T' this field is approximately equal to the meron field in Eq. (4.31), at least over distances small compared to $|T-T'|$. On the other hand, the 3D field at $x_4=0$ is not of long range but is of order $|\vec{x}|^{-2}$ as $|\vec{x}|\to\infty$, like a 3D dipole field.

The meron-pair configuration is related to the instanton solution of the SU(2) gauge theory.[13] In particular, both have Pontryagin index $\int d^4x F\tilde{F}=1$. The meron pair can be regarded as a deformation of the instanton field.

The temporal-gauge form of a meron-pair configuration with meron centers c and c' not at the same point in three-space ($\vec{c}\neq\vec{c}'$) is not known to me.

The action of a meron-pair configuration depends only on the separation in four-space $(c-c')^2$. It can be shown that the action is[13,14]

$$S=\frac{3\pi^2}{g^2}\ln\frac{(\vec{c}-\vec{c}')^2+(c_4-c_4')^2}{a^2}+S_c\,, \tag{4.34}$$

where a is a small length that arises either from smearing out the point singularities of the field at c and c', or from the finite lattice spacing in a lattice theory; and S_c is finite in the limit $(c-c')^2/a^2\to\infty$. The interaction of merons is proportional to the logarithm of their separation. The action of a single meron, found by considering the limit $(c-c')^2\to\infty$, is divergent for an infinite-volume system, but only logarithmically. For comparison, the action of a monopole string, i.e., a loop field where the curve C is an infinite string, which creates an isolated 3D monopole, has action that diverges linearly in an infinite volume.

There are two obvious differences between the contributions to the path integral of monopole loops and of merons. In the first place a monopole loop of size l creates a pair of *point* monopoles in the 3D subspace separated by a distance of order l. In contrast, a meron pair with meron centers c and c' creates a pair of 3D monopoles separated by a distance of order $|\vec{c}-\vec{c}'|$ and which depend in addition on scale parameters c_4 and c_4'. The scale parameter determines how large an effect the monopole has on the order parameters M or W; if $c_4\ll R$ then a monopole acts like a point monopole, but if $c_4\gtrsim R$ its effect is reduced. As seen before in Sec. III, this additional scale parameter changes the nature of the monopole sector as compared to a system of point monopoles.

In the second place the action of a monopole loop of size l is proportional to l/a. In contrast, the action of the meron-pair field is proportional to $\ln|c-c'|/a$. Thus the Boltzmann distribution $\exp(-S)$ suppresses monopole loops exponentially in l, but suppresses meron pairs only as a power of $|c-c'|$. In addition, the entropies of these configurations are very different. A monopole loop has one degree of freedom for each point on the loop, giving an entropy proportional to l. In contrast, a meron pair has only one degree of freedom, the relative position of the meron centers, giving an entropy proportional to $\ln|c-c'|$.

The logarithmic interaction (4.34) between merons and their logarithmic entropy suggest that there is a phase transition in which meron pairs ionize,[13] in the manner of the vortex ionization of a 2D Coulomb gas of vortices. Meron ionization would imply a monopole-ionization transition in the effective 3D theory defined by the vacuum functional $P_3(A_a^i)$. At small g^2 the action is large and only meron pairs with $(c-c')^2$ of order a^2 contribute to the path integral; in the effective 3D theory monopoles are combined in pairs with separation in both space and scale parameter of order a. At large g^2 the suppression due to the action is small and meron separations can be large compared to a. In the 3D theory free monopoles occur, although there are probably interactions between monopoles that produce some correlation in separation and scale parameter.

The usual estimate of the phase-transition point $g^2=g_c^2$, obtained by equating action and entropy, is[14]

$$\frac{g_c^2}{4\pi}=\frac{9\pi}{16}\,.$$

This value is very different than the value of g^2 at which the transition from strong- to weak-coupling behavior is observed in Monte Carlo studies of SU(2) gauge theories,[18,25] namely the much lower value $\beta=4/g^2\simeq 2$. This naive treatment of meron ionization is not an adequate description of the transition. Nevertheless, it is possible that a better treatment of the meron contribution to the path integral, including complete identification of meron degrees of freedom and interactions, and of the influence of the fluctuations of the gauge fields away from the meron configuration, would imply monopole ionization in the effective 3D theory and describe the transition from strong to weak coupling. An attempt to provide a better treatment of the meron contribution has been described by Laughton.[38]

Monopole loops may also play a role in the transition. However, there are some reasons to think that merons are more important. For example, in the small-g^2 phase, the fluctuations of large meron separations are suppressed only by a power, whereas large monopole loops are suppressed exponentially. Also, these configurations should not be counted separately because that would in-

troduce some double counting.

A characteristic feature of an ionization transiration of field configurations into just monopoles plus teraction is that the specific heat C is nonsingular at the transition point, as described earlier. For example, C is nonsingular for the ionization transition of a 2D Coulomb gas [see Eq. (2.35)], which supposedly explains the behavior of C for the XY model (Fig. 1). In contrast, C diverges at the transition point of the monopole-loop system. The specific heat of an SU(2) gauge theory is found in Monte Carlo studies to be nonsingular (Fig. 2).

The 3D monopoles in the disordered phase $g^2 > g_c^2$, can explain the area-law behavior of the loop integral W. However, for g^2 large compared to the transition value g_c^2, the strong-coupling expansion is an accurate approximation. The separation of field configurations into just monopoles plus small fluctuations is probably not adequate in that range of g^2, for a quantitative description of the area dependence of W.

Finally, the real problem of quark confinement in non-Abelian gauge theories is to understand the cause of the area-law behavior of W in the small-g^2 region, $g^2 < g_c^2$. The strong-coupling expansion cannot be used because the expansion is not convergent for exactly this range of g^2. Indeed the strong-coupling expansion is *expected* to break down for g^{-2} beyond a radius of convergence at which any kind of transition occurs. Also, since the 3D monopoles are bound in dipole pairs for $g^2 < g_c^2$, the contribution of the monopole sector to W is expected to give a perimeter law.

The crucial difference between non-Abelian and Abelian gauge theories is that the small fluctuations of the field $A_a^i(\vec{x}, x_4)$ away from the classical topological field configurations that minimize the action do not decouple in the non-Abelian case, but interact with the classical field. Physically this corresponds to the fact that gluons carry color charge while photons carry no electric charge. Thus the factorization into monopoles and small fluctuations does not occur as in Eq. (4.19) in a non-Abelian theory. Instead, the meron contribution to the partition function is schematically of the form

$$Z = \sum_n \frac{1}{n!} \sum_{c_1,\ldots,c_n} e^{-S_{\rm eff}(\bar{A})},$$

where $c_1, \ldots, c_n$ are meron center positions defining the classical configuration $\bar{A}$, and the effective action $S_{\rm eff}(\bar{A})$ is

$$e^{-S_{\rm eff}(\bar{A})} = \int d\varphi\, e^{-S(\bar{A}+\varphi)}.$$

The nonlinearity of S implies that $S_{\rm eff}(\bar{A})$ is not just $S(\bar{A})$ as it would be in an Abelian theory.

A possible explanation of the area-law behavior of $W(\Sigma)$ when g is less than g_c is as follows. The gluon fluctuations renormalize the charge by the replacement of $S(\bar{A})$ by $S_{\rm eff}(\bar{A})$. It may be that the effective monopole charge is a running charge $\kappa = 2\pi/g(R)$ that depends on the size R of Σ in the calculation of $W(\Sigma)$. Then even if the bare charge g is in the range $g < g_c$, the effective charge $g(R)$ would increase as the scale $R \to \infty$, as expected by asymptotic freedom, and become greater than g_c for R sufficiently large. Then $W(\Sigma)$ would see free monopoles and obey an area law. Further work is needed to find out whether this mechanism does operate.

V. THE XY-MODEL ANALOGY

In this final section the classical XY model is discussed in a way that emphasizes analogies with earlier discussions of gauge theories, particularly the topological order-disorder transition.

The XY model describes the statistical mechanics of classical two-component spins in a 2D lattice space with nearest-neighbor interactions. Lattice sites are denoted $\vec{x} = (x_1, x_2)$ and the lattice spacing is a. At each site there is a spin $\vec{s}(\vec{x})$ with fixed magnitude $|\vec{s}| = 1$; the spins can be specified by an angle variable $\theta(\vec{x})$ as

$$\vec{s}(\vec{x}) = (s_1(\vec{x}), s_2(\vec{x})) = (\cos\theta(\vec{x}),\ \sin\theta(\vec{x})). \tag{5.1}$$

The energy is

$$H = -J \sum_{\langle x, x'\rangle} \vec{s}(\vec{x}) \cdot \vec{s}(\vec{x}'), \tag{5.2}$$

where J is a coupling parameter and $\langle x, x'\rangle$ stands for nearest-neighbor pairs; each pair is counted once. The partition function is

$$Z = \int d\theta(\vec{x}) e^{-S(\theta)}, \tag{5.3}$$

where the action is

$$S(\theta) = K \sum_{\langle x, x'\rangle} \{1 - \cos[\theta(\vec{x}) - \theta(\vec{x}')]\} \tag{5.4}$$

and $K = \beta J$; the functional integral notation $\int d\theta(\vec{x})$ denotes $\prod_{\vec{x}} \int_{-\pi}^{\pi} d\theta(\vec{x})/2\pi$.

A quantity of special interest is the correlation function $A(R)$ defined by

$$A(R) = Z^{-1} \int d\theta(\vec{x}) e^{-S(\theta)} e^{i\theta(\vec{x})} e^{-i\theta(\vec{x}')}, \tag{5.5}$$

where $R = |\vec{x} - \vec{x}'|$. The importance of $A(R)$ is that its asymptotic behavior as $R \to \infty$ distinguishes between two phases of the model. If K is large, $A(R)$ decreases as a power of R as $R \to \infty$:

$$A(R) \sim c(a/R)^p \quad (K > K_c), \tag{5.6a}$$

where c is a constant and the exponent p depends on K; the phase characterized by this behavior is called critically ordered. If K is small, $A(R)$ decreases exponentially as $R\to\infty$:

$$A(R)\sim ce^{-\mu R/a}\quad (K<K_c), \tag{5.6b}$$

where c and μ are constants; this phase is called disordered. The phase transition separating these phases occurs at a critical value K_c of the parameter K.

The function $A(R)$ can be written

$$A(R)=Z^{-1}\int d\theta(\vec{x})e^{-S(\theta)}\exp\left(i\int_\Gamma d\vec{s}\cdot\nabla\theta\right), \tag{5.7}$$

where Γ is a path from $\vec{x}'$ to $\vec{x}$ and $R=|\vec{x}-\vec{x}'|$. The line-integral operator depends on $\theta(\vec{x})$ only along the one-dimensional (1D) subspace Γ. If Γ is linear then $A(R)$ is determined by an effective 1D theory. Let Γ be a segment of the x_1 axis. Define the reduced probability distribution $P_1(\varphi)$ of a 1D angular field $\varphi(x)$ by

$$P_1(\varphi)=Z^{-1}\int d\theta(\vec{x})e^{-S(\theta)}\prod_x\delta[\varphi(x)-\theta(x;0)]. \tag{5.8}$$

The distribution $P_1(\varphi)$ defines an effective 1D theory in which the partition function is

$$Z=\int d\varphi(x)P_1(\varphi) \tag{5.9}$$

and the correlation function $A(R)$ is

$$A(R)=Z^{-1}\int d\varphi(x)P_1(\varphi)\exp\left(i\int_\Gamma dx\,d\varphi/dx\right). \tag{5.10}$$

This reduction of dimensions to an effective 1D theory is analogous to the reduction to a 2D theory in the gauge theory [compare Eq. (2.1)]; the function $A(R)$ is the analog of the Wilson loop integral W [compare Eq. (2.2)].

To give a discussion of the XY model parallel to the earlier discussions of gauge theories, it is necessary to identify the configuration of the 1D field $\varphi(x)$ that is the analog of the 2D point vortex discussed in Sec. II. The relevant field is a point *kink* configuration defined by[39]

$$\varphi_k(x)=\pi\theta(x), \tag{5.11}$$

where $\theta(x)$ is the conventional θ function. The effect of a kink on the line integral in Eq. (5.10) is

$$\exp\left(i\int_\Gamma dx\,d\varphi_k/dx\right)=e^{i\pi\theta_\Gamma(0)}, \tag{5.12}$$

where the function $\theta_\Gamma(x)$ is defined by

$$\theta_\Gamma(x)=\begin{cases}1 & \text{if } x\in\Gamma,\\ 0 & \text{if } x\notin\Gamma.\end{cases} \tag{5.13}$$

The special significance of the kink field is that it makes $\exp(i\int dx\,d\varphi/dx)=-1$ and that it is of long range and so contributes equally for arbitrarily large Γ. Thus the kink field is the analog of the long-range vortex field with flux π/g in Eq. (2.4). An antikink configuration can be defined as $-\varphi_k(x)$ and multikink configurations by superposition; the general configuration is

$$\bar{\varphi}(x)=\sum_{i=1}^{n}\varphi_k(x-x_i)-\sum_{j=1}^{n'}\varphi_k(x-y_j), \tag{5.14}$$

where x_i and y_j are the positions of n kinks and n' antikinks. The line integral of $\bar{\varphi}$ is

$$\exp\left(i\int_\Gamma dx\,d\bar{\varphi}/dx\right)=\prod_{i,j}e^{i\pi\theta_\Gamma(x_i)}e^{-i\pi\theta_\Gamma(y_j)}. \tag{5.15}$$

A heuristic analysis of the effect of kinks on $A(R)$, in the spirit of the analysis of the effect of vortices on W in Sec. II, can be carried out by considering models of a gas of kinks, in which the functional integral $\int d\varphi(x)$ in Eqs. (5.9) and (5.10) is replaced by integrals over just the kink coordinates. Z and $A(R)$ become

$$Z=\sum_{n,n'}\frac{1}{n!n'!}\int\prod_{i,j}dx_i dy_j P(x,y), \tag{5.16}$$

$$A(R)=Z^{-1}\sum_{n,n'}\frac{1}{n!n'!}\times\int\prod_{i,j}dx_i dy_j P(x,y)e^{i\pi\theta_\Gamma(x_i)}e^{-i\pi\theta_\Gamma(y_j)}. \tag{5.17}$$

Simple models of the distribution $P(x,y)$ are considered.

For a perfect gas of kinks, i.e., the choice $P(x,y)=\zeta^{n+n'}$ where ζ is the density, the partition function is

$$Z=e^{2\zeta\Omega}, \tag{5.18a}$$

where $\Omega=\int dx$ is the volume of the 1D space. The correlation function is

$$A(R)=e^{-4\zeta R}, \tag{5.18b}$$

where R is the length of Γ. A perfect gas of kinks yields the exponential decrease of $A(R)$ as $R\to\infty$ associated with the disordered phase.

For a perfect gas of $k\bar{k}$ bound pairs, with members of each pair separated by l and density ζ', the partition function is

$$Z=\sum_n\frac{\zeta'^n}{n!}\prod_{i=1}^{n}\int dz_i\sum_{\sigma_i=\pm1}^{\frac{1}{2}}=e^{\zeta'\Omega}; \tag{5.19a}$$

here the degrees of freedom of a $k\bar{k}$ pair are the center of mass z_i and orientation σ_i which is $+1$ for $k\bar{k}$ or -1 for $\bar{k}k$. The correlation function $A(R)$ is, for $R\geq l$,

$$A(R)=e^{-4\xi/T}. \tag{5.19b}$$

A perfect gas of $k\bar{k}$ pairs yields a constant nonzero value of $A(R)$ as $R\to\infty$, as in an ordered phase.

The correlation function $A(R)$ is the analog of the loop integral W in the gauge theory. It is used to measure the disorder of the system. Further, the distinction between exponential decrease of $A(R)$ as $R\to\infty$ and slower decrease indicates the special topological disorder associated with uncorrelated 1D kink configurations. This suggests that the phase transition of the XY model is a topological order-disorder transition that has the effect of liberating kinks in the effective 1D theory. Thus the problem is to determine the origin of the kinks.

The nature of the phase transition of the XY model has been described by Kosterlitz and Thouless.[12] The starting point of their theory is the suggestion that the partition sum is adequately approximated if two kinds of configurations are included, long-range vortices in the spin field and small spin-wave fluctuations away from the classical vortex configurations. The phase transition, which occurs in the vortex sector of the system, is a topological order-disorder transition in which bound vortex-antivortex pairs dissociate.

The field of a point vortex centered at $\vec{x}=0$ is

$$\theta_v(\vec{x})=\tan^{-1}\frac{x_2}{x_1}. \tag{5.20}$$

The circulation of a contour C surrounding $\vec{x}=0$ in this field is

$$\oint_C d\vec{s}\cdot\nabla\theta_v=2\pi. \tag{5.21}$$

An antivortex field, with circulation -2π, is defined as $-\theta_v(\vec{x})$. Multivortex fields are obtained by superposition; the general configuration is

$$\bar{\theta}(\vec{x})=\sum_{i=1}^{n}\theta_v(\vec{x}-\vec{x}_i)-\sum_{j=1}^{n'}\theta_v(\vec{x}-\vec{y}_j). \tag{5.22}$$

The vortex field has a topological significance. The spin vector $\vec{s}(\vec{x})$ corresponding to $\theta_v(\vec{x})$ by Eq. (5.1) varies slowly over the entire lattice (except at the singular point $\vec{x}=0$); locally, the spins $\vec{s}(\vec{x})$ are almost aligned. But globally the spins vary over the entire range of directions; the direction of $\vec{s}(\vec{x})$ rotates through 2π as the path C is traversed. The quantization condition $\int_C d\vec{s}\cdot\nabla\theta=2\pi n$, as in Eq. (5.21), which implies that any discontinuity of $\theta_v(\vec{x})$ is a multiple of 2π, is imposed to avoid lines of misalignment of the spins.

Vortices of the spin field $\theta(\vec{x})$ create kinklike configurations in the effective 1D theory. The 1D field $\varphi(x)$ produced by a vortex centered at $\vec{c}$ is

$$\varphi(x)=\theta_v(x-c_1,\,-c_2)=\tan^{-1}\frac{-c_2}{x-c_1} \tag{5.23}$$

in accordance with Eq. (5.8). This field approaches 0 as $(x-c_1)\to\infty$ and π as $(x-c_1)\to-\infty$, like the field of an antikink. The parameter c_1 describes the position of the antikink. The additional parameter c_2 is a *scale* parameter that determines the size of the region over which φ changes by π. The 2D topological condition Eq. (5.21) ensures that the kink $\varphi(x)$ obeys the 1D condition $\varphi(\infty)-\varphi(-\infty)=\pi$.

Thus the vortex field in the XY model is the analog of the monopole field in the description of the gauge theory in Sec. III: each of these fields creates the topological fluctuations responsible for disorder in the effective theory of reduced dimension.

The contribution of vortices to the functional integral $\int d\theta(\vec{x})$ is obtained as usual by writing the integration variable $\theta(\vec{x})$ as

$$\theta(\vec{x})=\bar{\theta}(\vec{x})+\chi(\vec{x}), \tag{5.24}$$

where $\bar{\theta}(\vec{x})$ is the multivortex field (5.22) and $\chi(\vec{x})$ is a small spin-wave fluctuation, replacing the functional integral $\int d\theta(\vec{x})$ by integrals over the coordinates x_i and y_j of $\bar{\theta}$ and the spin-wave fluctuation $\chi(\vec{x})$, and treating $\chi(\vec{x})$ in the Gaussian approximation. The multivortex field $\bar{\theta}(\vec{x})$ is a solution of the field equations, except at the singular points of vorticity, so the action becomes $S(\bar{\theta}+\chi)\simeq S(\bar{\theta})+S(\chi)$ for small $\chi(\vec{x})$. The vortices and spin waves do not interact in the first approximation, so the partition function Z and correlation function $A(R)$ factorize into separate terms from the vortex and spin-wave sectors: $Z=Z_{sw}Z_v$ and $A=A_{sw}A_v$. The vortex sector has the partition function

$$Z_v=\sum_{n+n'}\frac{1}{n!\,n'!}\left(\frac{1}{a^2}\right)^{n+n'}\int\prod_{i,j}d^2x_i d^2y_j e^{-S(\bar{\theta})}, \tag{5.25}$$

where x_i and y_j are the v and v coordinates of the multivortex configuration $\bar{\theta}(\vec{x})$.

The spin-wave sector prevents long-range order in the large-K phase of the system: $A_{sw}(R)$ has the power-law dependence on R in Eq. (5.6a) for all K. However, it is the vortex sector that produces the phase transition to the disordered small-K phase, i.e., the exponential decrease of $A(R)$

in Eq. (5.6b).

The vortex sector of the model is equivalent to a 2D Coulomb gas of vortices. The action $S(\bar{\theta})$ is

$$S(\bar{\theta}) = \tfrac{1}{2}K \sum_{\lambda,\lambda'} e_\lambda e_{\lambda'} G(\vec{x}_\lambda - \vec{x}_{\lambda'}), \tag{5.26}$$

where $G(\vec{x})$ is the lattice Green's function Eq. (2.26) and $(e_\lambda, \vec{x}_\lambda)$ has the same meaning as in Eq. (2.25). Comparison of $S(\bar{\theta})$ with the Hamiltonian of the 2D Coulomb gas used in Sec. II [Eq. (2.24)] leads to the identification $K = \beta\kappa^2$. The density parameter ξ is just $1/a^2$, but in what follows it will be left arbitrary.

The expression (5.26) is only valid if $n = n'$. If $n \neq n'$ there is an additional term proportional to $(n-n')^2 \ln\Omega/a^2$ related to the divergent self-energy of a vortex; in the infinite-volume limit $\Omega/a^2 \to \infty$ this term forces the number of vortices and antivortices to be equal. However, in the ordered phase only $v\bar{v}$ bound pairs occur so the constraint $n = n'$ is satisfied dynamically; and in the disordered phase the v's and $\bar{v}$'s are relatively uncorrelated so the precise equality of n and n' is not important. The requirement $n = n'$ can be ignored.

The 2D Coulomb gas has already been discussed in Sec. II. There are two phases, an ordered phase for $K \geq K_c$ in which vortices are bound in $v\bar{v}$ pairs, and a disordered phase for $K \leq K_c$ in which free vortices occur. The phase transition at $K = K_c$ is due to ionization of $v\bar{v}$ pairs. The ionization is responsible for the behavior of the specific heat C observed in a Monte Carlo study of the XY model,[26] shown in Fig. 1. As β decreases, the fluctuation of vortex-antivortex separation increases and so the specific heat C increases; but C remains nonsingular at the transition point [see Eq. (2.35)] because the $v\bar{v}$ potential is logarithmic.

In Sec. II the vortex ionization was described in terms of its effect on the loop integral W defined in Eq. (2.11). In the XY model the quantity W is the expectation value of operator $\exp(\frac{1}{2}i \oint_{\partial\Sigma} d\vec{s} \cdot \nabla\bar{\theta})$ involving the circulation of $\partial\Sigma$. Thus W is more directly related to the point vortices than is $A(R)$, which depends only on the fluctuations in the effective 1D theory. In this way W is, for the XY model, the analog of the quantity M defined in Eq. (3.20) for the monopole sector of a gauge theory. Also, W and A_v are similar: the line integral that defines W measures the change of $\bar{\theta}$ around the closed curve $\partial\Sigma$ whereas A_v measures the change of $\bar{\theta}$ along the open segment Γ. The difference between an open or closed curve can be important even in the limit $R \to \infty$ because of the long-range nature of vortices. Either W or A_v can be used to identify topological disorder in the XY model, just as either M or W can be used in the gauge theory.

The contribution of the vortex sector to the correlation function $A(R)$ is

$$A_v = Z_v^{-1} \sum_{n,n'} \frac{1}{n!n'!} \prod_{i,j} \int d^2x_i d^2y_j P(x,y) \times e^{i\omega_\Gamma(x_i)} e^{-i\omega_\Gamma(y_j)}, \tag{5.27}$$

where $\omega_\Gamma(\vec{x})$, which is defined by

$$\omega_\Gamma(\vec{x}) = \int_\Gamma d\vec{s} \cdot \nabla\theta_v(\vec{s} - \vec{x}), \tag{5.28}$$

is the angle subtended by the line segment Γ at the point $\vec{x}$. The distribution $P(x,y)$ is the Boltzmann distribution $e^{-S(\bar{\theta})}$.

Before considering A_v for a 2D Coulomb gas, it is interesting for comparison purposes to look once again at some simpler models. Consider first the perfect gas of vortices, $P(x,y) = \xi^{n+n'}$; for this system A_v is

$$A_v = \exp\left\{-2\xi \int d^2x [1 - \cos\omega_\Gamma(\vec{x})]\right\}. \tag{5.29}$$

The integral diverges logarithmically; this *infrared* divergence is due to the slow decrease of $\omega_\Gamma(\vec{x})$ as $|\vec{x}| \to \infty$:

$$\omega_\Gamma(\vec{x}) \sim (R/r)\sin\varphi \quad \text{for } r \gg R, \tag{5.30}$$

where (r, φ) are the polar coordinates of $\vec{x}$ and R = length of Γ. It follows that A_v is of the form

$$A_v = \exp\left(-\xi \frac{\pi}{2} R^2 \ln\frac{\Omega}{R^2}\right), \tag{5.31}$$

where $\Omega = \int d^2x$. The disorder of a perfect gas of vortices make A_v vanish in the infinite-volume limit for any nonzero R. The kinks in the effective 1D theory corresponding to this system are completely uncorrelated in both position and scale. If only kinks of scale $\lesssim R$ contributed to A_v then the effective 1D density of kinks would be $\zeta = \xi R$ and A_v would have the behavior $A_v \sim \exp(-\xi R^2)$. However, the contribution due to kinks with scale $> R$ is sufficiently large to produce the $\ln\Omega$ infinity.

Consider next a perfect gas of $v\bar{v}$ pairs with separation l. The partition function is given in Eq. (2.16). The correlation function is

$$A_v = \exp\left\{-\xi' \int d^2z \int \frac{d\theta}{2\pi} [1 - \cos[\omega_\Gamma(\vec{x}) - \omega_\Gamma(\vec{x}')]]\right\} \tag{5.32}$$

with $\vec{x}, \vec{x}'$ the $v, \bar{v}$ positions in Eq. (2.18). The integrand is small unless exactly one of $\vec{x}$ and $\vec{x}'$ is near Γ. In the limit $R \gg l$ this only occurs if the $v\bar{v}$ pair is located within a distance l of one of the end points of Γ. This leads to the estimate

$$A_v \sim \exp(-\xi' 4\pi l^2) \quad \text{for } R \gg l. \tag{5.33}$$

Because the $v\bar{v}$ pair field is of short range, it does not produce the topological disorder needed to make $A_v \to 0$ as $R \to \infty$. The corresponding kinks in the effective 1D theory are strongly correlated in scale and position.

A_v remains nonzero as $R \to \infty$ for this system, whereas the loop integral W tends to zero as $\exp(-L(\partial\Sigma))$ [Eq. (2.20)]. The difference is that the contribution of a $v\bar{v}$ pair to the line integral along Γ is significant only if the pair is near an end point, but it contributes to the integral around $\partial\Sigma$ if the pair is anywhere along $\partial\Sigma$.

Finally, A_v must be considered for the 2D Coulomb gas in the approximations used previously in Sec. II. In the ordered phase for large K, the partition sum is dominated by configurations in which vortices are combined in $v\bar{v}$ pairs and interactions between vortices in different pairs are ignored. Then A_v becomes

$$A_v = \exp\left\{-\xi^2 \int d^2x\, d^2y\, e^{\beta\kappa^2 G(\vec{x}-\vec{y})} \times[1-\cos(\omega_\Gamma(x)-\omega_\Gamma(y))]\right\}, \tag{5.34}$$

where the replacement $K=\beta\kappa^2$ has been made. This should be compared to Eq. (2.38) for the loop integral W in the same approximation. The expression (5.34) is simplified by use of the approximate form (2.28) of $G(x)$ and the limiting form (5.33) of the integral over position and orientation of the $v\bar{v}$ pair at $\vec{x}, \vec{y}$; in the limit $R \to \infty$,

$$A_v \sim \exp\left\{-\xi^2 8\pi^2 \int_{r_0}^{\infty} dl \exp\left(-\frac{\beta\kappa^2}{2\pi}\ln\frac{l}{r_0}\right) l^3\right\} = \exp\left\{-4\pi^2\xi^2 r_0^4\left(\frac{\beta\kappa^2}{4\pi}-2\right)^{-1}\right\}. \tag{5.35}$$

Thus A_v approaches a nonzero constant as $R \to \infty$, showing the expected topological order. Note, however, that $A_v(R=\infty)$ tends to zero as $\beta\kappa^2/4\pi \to 2$. This is an indication of the phase transition at $\beta\kappa^2/4\pi = 2$. To be specific, the transition from the critically ordered to the disordered phase is identified with the change in the asymptotic form of $A(R)$ from power-law decrease to exponential decrease. The power law $A(R) \sim R^{-\eta}$ comes from the spin-wave factor $A_{sw}(R)$ for all $\beta\kappa^2$. The vortex factor $A_v(R)$ tends to a constant for $\beta\kappa^2/4\pi > 2$ and decreases exponentially for $\beta\kappa^2/4\pi < 2$. A sign of the transition is that $\lim_{R\to\infty} A_v(R)$ tends to zero as $\beta\kappa^2/4\pi \to 2$.

In the disordered phase for small $K=\beta\kappa^2$, the method used to calculate A_v is to transform the theory to the equivalent sine-Gordon theory. The sine-Gordon form of A_v is derived to be [compare Eqs. (2.44) and (2.45)]

$$A_v = \int d\varphi(\vec{x}) e^{-S_\Gamma(\varphi)} \Big/ \int d\varphi(\vec{x}) e^{-S(\varphi)}, \tag{5.36a}$$

where the action $S_\Gamma(\varphi)$ is

$$S_\Gamma(\varphi) = \int d^2x \left\{\frac{1}{2\beta\kappa^2}(\nabla\varphi)^2 + 2\xi[1-\cos(\varphi+\omega_\Gamma)]\right\}. \tag{5.36b}$$

The analysis of Eq. (5.36) proceeds in analogy with the analysis in Sec. II of the sine-Gordon form of W. In the first approximation A_v is estimated to be simply [compare Eq. (2.49)]

$$A_v \simeq e^{-S_\Gamma(\varphi_c)}, \tag{5.37}$$

where φ_c is the field that minimizes S_Γ and the factor $\tilde{A}$ coming from the small fluctuations of φ away from φ_c is ignored.

The field φ_c is not known to me. It can be analyzed qualitatively in the manner of the discussion of Eqs. (3.48)–(3.53). Briefly, φ_c is approximately equal to $\omega_\Gamma(\vec{x})$ where $\omega_\Gamma(\vec{x})$ is smooth; but $\omega_\Gamma(\vec{x})$ is discontinuous across Γ:

$$\lim_{\epsilon\to 0}[\omega_\Gamma(x_1,\epsilon)-\omega_\Gamma(x_1,-\epsilon)] = 2\pi\theta_\Gamma(x_1),$$

and for φ_c this discontinuity is smoothed out over a distance comparable to the Debye length $\lambda_D = (\xi\beta\kappa^2)^{-1/2}$. It follows that as R, the length of Γ, tends to infinity, $S_\Gamma(\varphi_c)$ is asymptotically

$$S_\Gamma(\varphi_c) = C\xi\lambda_D R, \tag{5.38}$$

where C is a numerical constant. Therefore, A_v decreases exponentially as $R \to \infty$, as anticipated for the low-β phase of the XY model.

The quantities A_v and W are similar in a system of vortices, as described earlier. In particular, both A_v and W decrease as $\exp(-R)$ where R is the length of Γ and $\partial\Sigma$, in the 2D vortex plasma. The similarity of these functions is analogous to the similarity of W and M in the 3D monopole models.

The analogies between the vortex sector of the XY model and the monopole sector of a gauge theory are summarized in Table I.

The final topic to be addressed is the definition of dual parameters of the order parameters $A(R)$ and W.

The dual parameters for the monopole sector of a gauge theory were defined in Sec. III as operators that create magnetic flux in a 3D subspace. The analogous dual parameters of the XY model are operators that create vorticity in a 1D subspace. Let $U(\vec{c})$ be the operator that creates a 1D kink at c_1 with scale c_2; it may be written

TABLE I. Table of analogies.

	Gauge theory	XY model
Topological configurations		
reduced dimension	2D vortex 3D monopole	1D kink
full dimension	meron monopole line	vortex
	meron pair monopole loop	vortex-antivortex pair
Correlation functions		
reduced dimension	W	A
full dimension	M	W

$$U(\vec{c}) = \exp\left(i \int_{-\infty}^{\infty} dx_1 \theta_v(x_1 - c_1, -c_2)\delta/i\delta\theta(x_1, 0)\right), \tag{5.39}$$

where $\theta_v(\vec{x})$ is the classical vortex field and the spin field $\theta(x_1, 0)$ is thought of as a quantum operator. The action of $U(\vec{c})$ on any functional of $\theta(x_1, 0)$ is

$$U(\vec{c})F[\theta(x_1, 0)] = F[\theta(x_1, 0) + \theta_v(x_1 - c_1, -c_2)].$$

The commutation relation involving $U(\vec{c})$ and the line integral along Γ that defines A_v is

$$U^\dagger(\vec{c}) \exp\left(i \int_\Gamma d\vec{s} \cdot \nabla\theta\right) U(\vec{c}) = e^{i\omega_\Gamma(\vec{c})} \exp\left(i \int_\Gamma d\vec{s} \cdot \nabla\theta\right), \tag{5.40}$$

where the function $\omega_\Gamma(\vec{c})$ was defined in Eq. (5.28). In particular, if $c_2 = 0$ then $\omega_\Gamma(c_1, 0) = \pi\theta_\Gamma(c_1)$; the operators commute if $c_1 \notin \Gamma$ and anticommute if $c_1 \in \Gamma$. This commutation relation is analogous to Eq. (3.63).

The operator $U(c_1)$ creates a 1D point kink at c_1. An operator that creates a kink-antikink pair at the end points c_1 and c_1' of a segment $\tilde{\Gamma}$ of the x_1 axis is $U(\tilde{\Gamma}) = U(c_1)U^\dagger(c_1')$. The commutation relation involving this operator is

$$U^\dagger(\tilde{\Gamma}) \exp\left(i \int_\Gamma d\vec{s} \cdot \nabla\theta\right) U(\tilde{\Gamma}) = e^{i\pi n(\Gamma, \tilde{\Gamma})} \exp\left(i \int_\Gamma d\vec{s} \cdot \nabla\theta\right), \tag{5.41}$$

where

$$n(\Gamma, \tilde{\Gamma}) = \begin{cases} 1 & \text{if } \Gamma, \tilde{\Gamma} \text{ partially overlap,} \\ 0 & \text{otherwise.} \end{cases}$$

The dual operators $U(\vec{c})$ or $U(\tilde{\Gamma})$ can be used to examine the topological disorder of the system, according to arguments precisely like those given at the end of Sec. III for the dual parameters of the monopole sector of a gauge theory.

VI. SUMMARY

The thesis of this paper is that the functional integrals over field configurations of gauge theories can be approximated by keeping just two kinds of configurations: classical configurations of topological origin, and small Gaussian fluctuations. The topological configurations can be considered in various dimensions—vortices in two dimensions, magnetic monopoles in three dimensions, merons in four dimensions—and are the long-range fields that have a significant effect on the Wilson loop integral W. In this paper, explicit statistical-mechanics models of the contribution of these fields to the functional integrals are studied in order to try to understand their influence on the large-distance behavior of the Wilson loop integral W. This approach is meant to be heuristic. Further work is needed to make more precise the connection between the model calculations and the full field theory.

The results of these models show the importance of the interactions between 2D vortices or 3D monopoles. For example, in the disordered phase of the 2D Coulomb gas of vortices, the correlations of vortex positions responsible for Debye screening imply that W obeys a perimeter law, not an area law as would result if the screening were ignored. In addition, an additional dynamical coordinate, the *scale* of the configuration, arises when the configuration derives from one higher dimension; the scale parameter can significantly change the nature of the system, compared to a system of point fields.

In several of the models an ultraviolet momentum-space cutoff is needed to render the results

finite. The ultraviolet divergences are associated with point singularities of the vortex or monopole fields, and can be eliminated by making the monopoles have nonzero spatial extent. Alternatively, the ultraviolet cutoff is provided naturally in a lattice theory. However, the interesting questions concern the large-distance behavior of the system, and this may be insensitive to the method used to treat the ultraviolet singularities. In particular, the long-range nature of the topological configurations implies that they vary slowly and can be described by a continuum approximation.

Ionization of monopole pairs in the effective 3D theory may explain the transition from weak to strong coupling in gauge theories: At large $\beta = 1/g^2$ the system is ordered and only small fluctuations of the gauge fields, gluon fluctuations plus $m\bar{m}$ dipole pairs, occur. At small β the system is disordered; for $\beta = 0$ all field configurations are equally probable. Monopole ionization carries the system from large-β to small-β behavior. This suggestion might be studied by considering quantities like M [Eq. (3.20)] or the dual parameter $U(\vec{c})$ [Eq. (3.62)] that are directly related to monopole fields. Also, the Monte Carlo studies of gauge theories may provide a laboratory for testing models of this transition. The specific-heat curve in Fig. 2, for example, seems particularly interesting. The models considered here are not good enough to reproduce quantitatively the Monte Carlo results, but they do indicate some of the qualitative features.

If it is true that monopole ionization is the origin of the transition from weak to strong coupling, then calculation of W involves an interesting and complex interplay of ideas from topology, classical field theory, and statistical mechanics. The last aspect of the problem is emphasized in this paper. Other aspects, such as the problem of identifying all the degrees of freedom that specify a monopole configuration in a non-Abelian theory, are ignored.

Finally, the XY model is interesting as a simpler analog system in which ideas about topological order-disorder transitions can be tested.

ACKNOWLEDGMENTS

I am pleased to thank D. H. Weingarten, M. Creutz, and S. Shenker for conversations and suggestions about this work. I am especially indebted to D. N. Petcher for giving me access to the results of his Monte Carlo studies of SU(2) gauge theory prior to their publication. This work was supported by the U. S. Department of Energy.

[1]D. R. Stump, Phys. Rev. D 20, 1002 (1979).
[2]D. R. Stump, Phys. Rev. D 20, 1965 (1979).
[3]D. R. Stump, Phys. Rev. D 21, 2370 (1980).
[4]T. T. Wu and C. N. Yang, in *Properties of Matter Under Unusual Conditions*, edited by H. Mark and S. Fernbach (Interscience, New York, 1969).
[5]C. M. Bender, T. Eguchi, and H. Pagels, Phys. Rev. D 17, 1086 (1978); K. Cahill and D. R. Stump, *ibid.* 20, 540 (1979).
[6]S. Mandelstam, Phys. Rep. 23C, 245 (1976); Phys. Rev. D 19, 2391 (1979); in *Proceedings of the 1979 International Symposium on Lepton and Photon Interactions at High Energies, Fermilab*, edited by T. B. W. Kirk and H. D. I. Abarbanel (Fermilab, Batavia, Illinois, 1979).
[7]S. D. Drell, in the Goldhaber Festschrift (unpublished).
[8]K. G. Wilson, Phys. Rev. D 10, 2445 (1974).
[9]K. Cahill and D. R. Stump, Phys. Rev. D 20, 2096 (1979); L. S. Brown and W. I. Weisberger, *ibid.* 20, 3239 (1979).
[10]A. M. Polyakov, Nucl. Phys. B120, 429 (1977).
[11]C. G. Callan, R. F. Dashen, and D. J. Gross, Phys. Lett. 63B, 334 (1976).
[12]J. M. Kosterlitz and D. J. Thouless, J. Phys. C 6, 1181 (1973); J. M. Kosterlitz, *ibid.* 7, 1046 (1974).
[13]C. G. Callan, R. F. Dashen, and D. J. Gross, Phys. Lett. 66B, 375 (1977); Phys. Rev. D 17, 2717 (1978).
[14]D. R. Stump, Phys. Rev. D 20, 2592 (1979).
[15]R. Balian, J. Drouffe, C. Itzykson, Phys. Rev. D 11, 2098 (1975).
[16]A. M. Polyakov, Phys. Lett. 59B, 79 (1975); T. Banks, R. Myerson, and J. Kogut, Nucl. Phys. B129, 493 (1977); M. E. Peskin, Ann. Phys. (N.Y.) 113, 122 (1978); E. Fradkin and L. Susskind, Phys. Rev. D 17, 2637 (1978).
[17]J. Kogut, R. Pearson, and J. Shigemitsu, Phys. Rev. Lett. 43, 484 (1979).
[18]M. Creutz, Phys. Rev. Lett. 43, 553 (1979); Phys. Rev. D 21, 2308 (1980).
[19]S. Samuel, Nucl. Phys. B154, 62 (1979).
[20]C. G. Callan, R. F. Dashen, and D. J. Gross, Phys. Rev. Lett. 44, 435 (1980).
[21]G. 't Hooft, Nucl. Phys. B138, 1 (1978); B153, 141 (1979).
[22]Construction of non-Abelian vortex configurations is described by Samuel in Ref. 19.
[23]See, for example, P. Minnhagen, A. Rosengren, and G. Grinstein, Phys. Rev. B 18, 1356 (1978).
[24]M. Creutz, L. Jacobs, and C. Rebbi, Phys. Rev. Lett. 42, 1390 (1979); Phys. Rev. D 20, 1915 (1979); C. Rebbi, Brookhaven report (unpublished).
[25]D. N. Petcher and D. H. Weingarten, Phys. Rev. D 22, 2465 (1980).
[26]J. Tobochnik and G. V. Chester, Phys. Rev. B 20, 3761 (1979).
[27]V. L. Berezinskii, Zh. Eksp. Teor. Fiz. 59, 907 (1970) [Sov. Phys.-JETP 32, 493 (1971)]; J. V. Jose, L. P. Kadanoff, S. Kirkpatrick, and D. R. Nelson, Phys. Rev. B 16, 1217 (1977).

[28]For comparison with the notation of Ref. 12, the charge q^2 used there is $q^2 = \varkappa^2/4\pi$.
[29]See Eq. (26) of Ref. 12.
[30]A. Luther and I. Peschel, Phys. Rev. B 12, 3908 (1975); A. Luther and D. J. Scalapino, *ibid.* 16, 1153 (1977); P. B. Wiegmann, J. Phys. C 11, 1583 (1978); S. Coleman, Phys. Rev. D 11, 2088 (1975); see also B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979), Sec. 23 and references therein.
[31]L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd edition (Pergamon, New York, 1977), Secs. 48 and 70.
[32]G. 't Hooft, Nucl. Phys. B79, 276 (1974); A. M. Polyakov, Pis'ma Zh. Eksp. Teor. Fiz. 20, 430 (1974) [JETP Lett. 20, 194 (1974)].
[33]Here $\cos\Phi_D$ in Eq. (3.17) and $\exp(i\Phi_D)$ in Eq. (3.14) are equivalent because only the real part contributes.
[34]K. Huang, lectures at "Ettore Majorana", International School of Subnuclear Physics, Erice, 1978, MIT Report No. CTP 729 (unpublished).
[35]P. A. M. Dirac, Proc. R. Soc. London A133, 60 (1931); Phys. Rev. 74, 817 (1948).
[36]This estimate is also described by Banks *et al.* (Ref. 16).
[37]V. de Alfaro, S. Fubini, and G. Furlan, Phys. Lett. 65B, 163 (1976); 72B, 203 (1977).
[38]D. G. Laughton, Can. J. Phys. 58, 845 (1980); 58, 859 (1980).
[39]Perhaps this field should be called a half-kink, because the spins rotate by π along the x_1 axis, not 2π [Eq. (5.12)].